

Aus der Klinik und Poliklinik für Psychiatrie und Psychotherapie
der Ludwig-Maximilians-Universität München
Direktor: Univ. Prof. Dr. med. Peter G. Falkai

**Diagnostik positiver Antwortverzerrungen
bei forensischen Begutachtungen.
Untersuchung der Validität und Brauchbarkeit
der Supernormalität-Skala (SN-S)**

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von
Susanne Tippelt

aus
München

Jahr
2015

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter: Prof. Dr. Norbert Nedopil

Mitberichterstatter: Priv. Doz. Dr. Cornelis Stadtland
Priv. Doz. Dr. Johannes Levin

Dekan: Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 27.04.2015

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Tabellenverzeichnis	IV
Abbildungsverzeichnis	VII
Abkürzungsverzeichnis	VIII
Zusammenfassung	X
1 Einleitung	1
2 Theoretischer Hintergrund	2
2.1 Begriffsbestimmung Antwortverzerrungen	2
2.2 Positive Antwortverzerrungen im forensischen Begutachtungskontext	4
2.2.1 Häufigkeit von positiven Antwortverzerrungen bei forensischen Begutachtungen.....	4
2.2.2 Einflussfaktoren auf das Antwortverhalten.....	5
2.2.2.1 Untersuchungskontext bzw. -zweck	5
2.2.2.2 Probandenmerkmale	7
2.2.2.3 Untersuchungsmaterial	12
2.2.3 Auswirkung von Antwortverzerrungen auf testpsychologische Befunde.....	14
2.3 Methoden zur Erfassung von Antwortverzerrungen.....	16
2.4 Validierung von Verfahren zur Erfassung von Antwortverzerrungen.....	18
2.4.1 Validitätsindizes.....	18
2.4.2 Untersuchungsdesign	20
2.5 Die Supernormalität-Skala (SN-S)	23
3 Fragestellungen	31
4 Methoden	32
4.1 Die Stichprobe	32
4.1.1 Rekrutierung und Zusammensetzung der Kontrollstichprobe	32
4.1.2 Zusammensetzung der Stichprobe der Begutachtungsprobanden	34

4.1.3	Beschreibung der Stichprobe	35
4.2	Stichprobe zur Kreuzvalidierung (Bantel, 2011)	40
4.3	Untersuchungsdesign und Erhebungsinstrumente	40
4.3.1	Onlinefragebogen.....	41
4.3.2	Psychopathic Personality Inventory-Revised (PPI-R)	43
4.3.3	Persönlichkeits-Stil- und Störungsinventar (PSSI).....	44
4.3.4	Inventar Klinischer Persönlichkeitsakzentuierungen (IKP).....	45
4.3.5	Fragebogen zur Erfassung von Aggressivitätsfaktoren (FAF)	46
4.3.6	Rosenzweig Picture-Frustration Test (PFT)	47
4.3.7	Foto-Hand-Test (FHT).....	49
4.3.8	Skala „Impression Management“ aus dem 16 Persönlichkeitsfaktoren-Test – Revidierte Fassung (16 PF-R).....	50
4.3.9	Skala „Offenheit“ aus dem Eppendorfer Schizophrenie-Inventar (ESI)	51
4.3.10	Kontrollskalen des Minnesota Multiphasic Personality Inventory – 2 (MMPI-2).....	51
4.3.11	Strukturierter Fragebogen Simulierter Symptome (SFSS)	53
4.4	Statistische Analyse der Daten.....	54
5	Ergebnisse	58
5.1	Item-Statistiken und Reliabilitätsanalyse.....	58
5.2	Faktorenanalyse	61
5.3	Berechnung der Cut-off-Werte und diagnostische Validität	65
5.3.1	SN-S-Gesamtwert	65
5.3.2	SN-S-Subskalen	66
5.3.3	Aus der Faktorenanalyse extrahierte Skalen.....	69
5.3.4	PPI-Skalen.....	71
5.4	Konvergente Validität	72
5.4.1	Dimensionale Übereinstimmung mit anderen Kontrollskalen.....	72
5.4.2	Kategoriale Übereinstimmung mit anderen Kontrollskalen	74

5.5	Erfassung von negativen Antwortverzerrungen mit der SN-S.....	76
5.6	Die Häufigkeit des Auftretens von positiven Antwortverzerrungen	78
5.6.1	Mittelwertsunterschiede zwischen den verschiedenen Substichproben.....	78
5.6.2	Häufigkeitsunterschiede zwischen den verschiedenen Substichproben	80
5.7	Auswirkung von positiven Antwortverzerrungen auf testpsychologische Befunde	85
5.7.1	Aggressionsbereitschaft	85
5.7.2	Persönlichkeitsstörungen	90
5.7.2.1	Anzahl unterdurchschnittlicher Skalenwerte in klinischen Persönlichkeitsinventaren	90
5.7.2.2	Diskrepanz zwischen klinischer Einschätzung und Selbstbericht	91
6	Diskussion.....	92
6.1	Zusammenfassung und Diskussion der Ergebnisse	92
6.1.1	Reliabilitätsanalyse und Bildung von Subskalen	92
6.1.2	Cut-off-Werte und diagnostische Validität	93
6.1.3	Konvergente Validität	94
6.1.4	Erfassung von negativen Antwortverzerrungen mit der SN-S.....	96
6.1.5	Auftretenshäufigkeit von Supernormalität in Abhängigkeit vom Untersuchungskontext.....	97
6.1.6	Auswirkung von Supernormalität auf testpsychologische Befunde	98
6.2	Diskussion des methodischen Vorgehens	102
6.3	Fazit und Empfehlungen für die Praxis	106
7	Literaturverzeichnis	108
8	Anhang	121

Tabellenverzeichnis

Tabelle 1	Alter der untersuchten Probanden in Jahren.....	35
Tabelle 2	Demographische Merkmale der untersuchten Probanden (Häufigkeit, Prozentwerte in Klammern).....	36
Tabelle 3	Vom psychiatrischen Gutachter gestellte Hauptdiagnosen (ICD-10 F-Kategorien) und diagnostizierte Persönlichkeitsstörungen (PS) (Häufigkeit, Prozentwerte in Klammern).....	38
Tabelle 4	Indexdelikte der Strafrechtsprobanden (Häufigkeit, Prozentwerte in Klammern).....	39
Tabelle 5	Alter in Jahren und Geschlecht (Häufigkeit, Prozentwerte in Klammern) der Probanden von Bantel (2011) nach Stichprobenkorrektur	40
Tabelle 6	Interne Konsistenz (Cronbachs α) und mittlere Inter-Item-Korrelation (MIC) des SN-S-Gesamtwerts und der Subskalen in der Gesamtstichprobe (GS) und den einzelnen Substichproben.....	58
Tabelle 7	Interne Konsistenz (Cronbachs α) und mittlere Inter-Item-Korrelation (MIC) der PPI-R Skalen Unaufrichtige Beantwortung und Soziale Erwünschtheit in der Gesamtstichprobe (GS) und den einzelnen Substichproben	60
Tabelle 8	Faktorladungen der SN-S Items bei einer Hauptkomponentenanalyse mit Promax-Rotation.....	62
Tabelle 9	Interne Konsistenz (Cronbachs α) der aus der Faktorenanalyse extrahierten Skalen in der Gesamtstichprobe (GS) und den einzelnen Substichproben.....	65
Tabelle 10	Übereinstimmung der Klassifikation des Antwortverhaltens mit dem SN-S-Gesamtwert und der Untersuchungsgruppe	66
Tabelle 11	Übereinstimmung der Klassifikation des Antwortverhaltens durch die SN-S-Subskalen mit der Untersuchungsgruppe.....	67
Tabelle 12	Übereinstimmung der Klassifikation des Antwortverhaltens durch die aus der Faktorenanalyse extrahierten SN-S-Subskalen mit der Untersuchungsgruppe	69
Tabelle 13	Übereinstimmung der Klassifikation des Antwortverhaltens durch die aus der Faktorenanalyse extrahierten SN-S-Subskalen mit der Untersuchungsgruppe in der Kreuzvalidierungsstichprobe.....	70

Tabelle 14	Übereinstimmung der Klassifikation des Antwortverhaltens durch die PPI-UA-Skala mit der Untersuchungsgruppe.....	72
Tabelle 15	Übereinstimmung der Klassifikation des Antwortverhaltens durch die PPI-SE-Skala mit der Untersuchungsgruppe.....	72
Tabelle 16	Korrelationen des SN-S-Gesamtwerts (GW) und der Subskalen mit den anderen Kontrollskalen zur Erfassung positiver und negativer Antwortverzerrungen.....	73
Tabelle 17	Übereinstimmung der Klassifikation des Antwortverhaltens durch den SN-S-Gesamtwert und andere Kontrollskalen zur Erfassung positiver Antwortverzerrungen.....	74
Tabelle 18	Übereinstimmung der Klassifikation des Antwortverhaltens durch den SN-S-Gesamtwert und die MMPI-Kontrollskalen.....	76
Tabelle 19	Korrelationen des SN-S-Gesamtwerts (GW) und der SN-S-Subskalen mit SFSS-Gesamtwert (GW) und den SFSS-Subskalen.....	77
Tabelle 20	Übereinstimmung der Klassifikation des Antwortverhaltens durch den SN-S-Gesamtwert und den SFSS-Gesamtwert.....	78
Tabelle 21	Häufigkeit auffälliger Werte auf den einzelnen SN-S-Skalen in den verschiedenen Untersuchungsgruppen	82
Tabelle 22	Vergleich der Aggressionsbereitschaft (Mittelwerte, Standardabweichungen) im FAF (FAF-Agg T-Wert), dem Foto-Hand-Test (FHT-AOS) und dem PFT (PFT-E) zwischen Probanden mit unauffälligen (> Cut-off) und auffälligen (< Cut-off) SN-S-Werten.....	86
Tabelle 23	Vergleich der kategorialen Auswertung von FAF-Agg, FHT-AOS und PFT-E zwischen Probanden mit unauffälligen bzw. auffälligen SN-S-Ergebnissen.....	87
Tabelle 24	Diskrepanz zwischen selbstberichteter und mithilfe (semi-)projektiver Verfahren erfasster Aggressionsbereitschaft (Aggravation, Übereinstimmung, Dissimulation) in Abhängigkeit von den SN-S-Ergebnissen.....	89
Tabelle 25	Mittwerte (<i>M</i>), Standardabweichungen (<i>SD</i>) und korrigierte Trennschärfen (<i>r</i>) für den Gesamtwert der 34 SN-S-Items in der Gesamtstichprobe und den Substichproben.....	135

Tabelle 26	Mittwerte (<i>M</i>), Standardabweichungen (<i>SD</i>) und korrigierte Trennschärfen (<i>r</i>) für den Gesamtwert der 23 Items der PPI-Skala Unaufrichtige Beantwortung in der Gesamtstichprobe und den Substichproben.....	138
Tabelle 27	Mittwerte (<i>M</i>), Standardabweichungen (<i>SD</i>) und korrigierte Trennschärfen (<i>r</i>) für den Gesamtwert der 11 Items der PPI-Skala Soziale Erwünschtheit in der Gesamtstichprobe und den Substichproben	140
Tabelle 28	Übereinstimmung der Klassifikation des Antwortverhaltens durch SN-S-Subskalen und die Kontrollskalen des PPI-R	143
Tabelle 29	Deskriptive Statistiken des SN-S-Gesamtwerts und der Subskalen in der Gesamt- und den Substichproben	147

Abbildungsverzeichnis

Abbildung 1	Verteilung des SN-S-Gesamtwerts in den Substichproben	79
Abbildung 2	Zusammenhang zwischen dem SN-S-Gesamtwert und den Werten auf den Kontrollskalen des PPI-R, des FAF, des 16 PF-R und des ESI.....	141
Abbildung 3	Zusammenhang zwischen dem SN-S-Gesamtwert und den Werten auf den Kontrollskalen des MMPI-2.....	142
Abbildung 4	Zusammenhang zwischen dem SFSS-Gesamtwert und dem Gesamtwert der SN-S sowie den Subskalen Affektive Symptome, Zwang, Dissoziative Symptome und Psychotische Symptome	145
Abbildung 5	Zusammenhang zwischen dem SFSS-Gesamtwert und den SN-S-Skalen Aggression, Minimierung depressiver Beschwerden, Minimierung von Aggression und Minimierung von ungewöhnlichem Erleben	146
Abbildung 6	Verteilung der Werte auf den SN-S-Skalen Soziale Erwünschtheit (4 Items), Affektive Symptome (7 Items) und Zwang (4 Items) in den Substichproben.....	149
Abbildung 7	Verteilung der Werte auf den SN-S-Skalen Dissoziatives Erleben (5 Items), Psychotische Symptome (6 Items) und Aggression (7 Items) in den Substichproben.....	150
Abbildung 8	Verteilung der Werte auf den SN-S-Skalen Minimierung depressiver Beschwerden (9 Items), Minimierung von Aggression (8 Items) und Minimierung von ungewöhnlichem Erleben (6 Items) in den Substichproben.....	151
Abbildung 9	Zusammenhang zwischen dem SN-S-Gesamtwert und der Anzahl unterdurchschnittlicher Skalen in IKP-G, IKP-Eg und PSSI	152

Abkürzungsverzeichnis

BtMG	Betäubungsmittelgesetz
ESI	Eppendorfer Schizophrenie-Inventar
-FR	Skala Offenheit
Fake-KG	Kontrollgruppe mit Fake-good-Instruktion
GN	Gesundheitsnachweis
GS	Gesamtstichprobe
GW	Gesamtwert
IKP	Inventar Klinischer Persönlichkeitsakzentuierungen
-G	Grundinventar
-Eg	Ergänzungsmodul
KG	Kontrollgruppe
KV	Kreuzvalidierung
MMPI-2	Minnesota Multiphasic Personality Inventory-2
-F	Seltenheitsskala
-L	Lügenskala
-K	Korrekturskala
FAF	Fragebogen zur Erfassung von Aggressivitätsfaktoren
-Agg	Gesamtwert „Σ Aggressivität“
-Off	Skala Offenheit
FHT-AOS	Acting-Out-Score aus dem Foto-Hand-Test
PSSI	Persönlichkeits-Stil- und Störungs-Inventar
PFT-E	Anzahl der aggressiven Antworten im Rosenzweig Picture-Frustration Test
RP	Rückfallprognose
SF	Schuldfähigkeit
SFSS	Strukturierter Fragebogen Simulierter Symptome
-AM	Skala Amnestische Störungen
-AS	Skala Affektive Störungen
-N	Skala Neurologische Beeinträchtigung
-NI	Skala Niedrige Intelligenz
-P	Skala Psychose

SN-S	Supernormalität-Skala
-AGG	Skala Aggression
-AS	Skala Affektive Symptome
-DE	Skala Dissoziative Symptome
-MDe	Skala Minimierung depressiver Beschwerden
-MA	Skala Minimierung von Aggression
-MUE	Skala Minimierung von ungewöhnlichem Erleben
-PS	Skala Psychotische Symptome
-SE	Skala Soziale Erwünschtheit
-ZW	Skala Zwang
StGB	Strafgesetzbuch
PPI-R	Psychopathic Personality Inventory-Revised
-SE	Skala Soziale Erwünschtheit
-UA	Skala Unaufrichtige Beantwortung
ZB	Zahlungsbegehren
16PF-R	16 Persönlichkeitsfaktoren-Test – Revidierte Fassung (16 PF-R)
-IM	Skala Impression Management

Zusammenfassung

Bei forensischen Begutachtungen können testpsychologische Untersuchungen fragestellungsrelevanter Persönlichkeitsmerkmale nur dann einen sinnvollen Beitrag zur Gesamtbeurteilung liefern, wenn die Ergebnisse valide interpretiert werden können. Da die Befunde von den Probanden durch Antwortverzerrungen verfälscht werden können, bedarf es bei jeder Begutachtung einer multimodalen Diagnostik des Antwortverhaltens. Allerdings sind für positive Antwortverzerrungen, d.h. das Herunterspielen und Leugnen von Auffälligkeiten und Beschwerden, bislang kaum validierte Testverfahren verfügbar. Die Supernormalität-Skala (SN-S) wurde entwickelt, um diese Lücke zu schließen. Ziel der vorliegenden Arbeit war eine umfassende empirische Überprüfung der autorisierten deutschen Übersetzung.

Die diagnostische Validität wurde mit einem experimentellen Design anhand einer heterogenen Kontrollstichprobe ($N = 201$) mithilfe eines Onlinefragebogens mit Standardinstruktion und Fake-Instruktion bestimmt. Für die Überprüfung weiterer Aspekte der Kriteriums- und der Konstruktvalidität, insbesondere der Frage, ob die mit der SN-S erfassten Antwortverzerrungen Auswirkungen auf (test-)psychologische Befunde haben, wurden zudem Daten von 108 Begutachtungsprobanden analysiert.

Die SN-S zeigte eine ausreichende diagnostische Validität ($AUC = .86$), wobei insbesondere die Sensitivität (94%) und der negative Vorhersagewert (96%) hoch waren, Spezifität (69%) und positiver Vorhersagewert (62%) waren dagegen nicht zufriedenstellend. Damit bedürfen auffällige SN-S-Werte einer weiteren Überprüfung mit anderen Verfahren. Sowohl die Subskalen der Originalversion als auch die mithilfe einer Faktorenanalyse extrahierten Skalen wiesen eine deutlich schlechtere Klassifikationsgüte auf. Überwiegend moderate bis hohe Korrelationen ($r = .24-.80$) mit anderen Skalen zur Erfassung positiver und negativer Antwortverzerrungen zeigen, dass Supernormalität kein eindimensionales Konstrukt ist; mit der SN-S werden soziale Erwünschtheit und Dissimulationstendenzen, aber auch Aggravation und Simulation erfasst. Je nach Fragestellung der Begutachtung (Schuldfähigkeit, Rückfallprognose, sozial- und zivilrechtliche Fragestellungen) zeigten zwischen 38% und 88% der Probanden supernormales Antwortverhalten. Probanden mit auffälligem SN-S-Ergebnis berichteten gegenüber Probanden mit unauffälligem Antwortverhalten eine signifikant niedrigere Aggressionsbereitschaft. Sie minimierten auch weitere klinische Persönlichkeitsmerkmale. Aus diesem Grund ist davon auszugehen, dass Fragebögen zur Er-

fassung von Persönlichkeitsstörungen bei Personen mit Antwortverzerrungen im forensischen Kontext nicht zu validen Ergebnissen führen.

Die SN-S ist ein brauchbares Instrument zur Überprüfung positiver Antwortverzerrungen bei forensischen Begutachtungen, als alleiniger Indikator aber noch unzureichend, um Disimulation zu belegen.

1 Einleitung

Der Einsatz von Selbstbeurteilungsverfahren bei forensischen Begutachtungen ist umstritten, da die Ergebnisse von Verfahren, die ausschließlich auf Auskünften der Probanden beruhen, bewusst verfälscht werden können, indem Probleme und Auffälligkeiten überbetont oder vorgetäuscht (negative Antwortverzerrungen) bzw. heruntergespielt oder geleugnet (positive Antwortverzerrungen) werden (Heilbrun, 1992; Heilbrun, Marczyk, DeMatteo & Mack-Allen, 2007; Kröber, 2005; Otto, 2008). Um im individuellen Fall zu entscheiden, ob der Einsatz von Fragebögen bei einer Begutachtung sinnvoll ist bzw. ob die Selbstauskünfte eines Probanden zuverlässig interpretiert werden können oder aufgrund von Täuschungs- und Manipulationsversuchen verzerrt sind, bedarf es einer Einschätzung bzw. Überprüfung des Antwortverhaltens. Die Diagnostik von Antwortverzerrungen ist daher unverzichtbar bei testpsychologischen Untersuchungen. Heilbrun et al. (2007) empfehlen in ihren Richtlinien für forensisch-psychiatrische bzw. -psychologische Begutachtungen die Überprüfung des Antwortverhaltens zum einen durch den Abgleich der Angaben des Probanden mit Informationen aus Akten oder Fremdanamnesen, d.h. eine Diskrepanzdiagnostik (Steller, 1994). Die Richtlinien sehen zum anderen vor, dass bei Verdacht auf das Vorliegen von Antwortverzerrungen zur weiteren Überprüfung spezifische Testverfahren bzw. Skalen eingesetzt werden. Auf das Antwortverhalten eines Probanden soll im Gutachten stets explizit eingegangen werden.

Für negative Antwortverzerrungen wurde eine Reihe von Verfahren entwickelt und validiert (Merten, 2014; Merten, Stevens & Blaskewitz, 2009), für die Erfassung positiver Antwortverzerrungen stehen dem Untersucher dagegen vor allem im deutschen Sprachraum bislang kaum geeignete Instrumente zur Verfügung. Deshalb wurde in der Abteilung für Forensische Psychiatrie der Ludwig-Maximilians-Universität München die von einer niederländischen Arbeitsgruppe entwickelte Supernormalität-Skala, die Dissimulationstendenzen erfasst (Cima, van Bergen & Kremer, 2008), ins Deutsche übersetzt (Veress, 2010) und ihre Validität in kleineren Untersuchungen überprüft (Bantel, 2011; Veress, 2010; Tippelt, Yundina, Nedopil & Veress, 2011; Zimmermann, 2012). Aufgrund erster vielversprechender Ergebnisse wurde das Instrument bei vielen Begutachtungen eingesetzt. Die diagnostische Validität des Fragebogens, d.h. die Güte bei der Klassifikation des Antwortverhaltens als verzerrt oder unauffällig, wurde bislang jedoch nicht empirisch überprüft. Dies war Ziel der vorliegenden Arbeit. Zudem sollten weitere Aspekte der Kriteriumsvalidität und der Konstruktvalidität bzw. die Brauchbarkeit der Supernormalität-Skala für die

Erfassung von positiven Antwortverzerrungen bei forensischen Begutachtungen überprüft werden, indem explorativ die Auswirkungen von „Supernormalität“ auf testpsychologische Ergebnisse untersucht werden. Denn nur wenn gezeigt werden kann, dass die mit der SN-S erfassten Antwortverzerrungen Auswirkungen auf (test-)psychologische Befunde haben, also die Ergebnisse der SN-S praktische Relevanz besitzen, ist ein weiterer Einsatz bei forensischen Begutachtungen gerechtfertigt.

2 Theoretischer Hintergrund

2.1 Begriffsbestimmung Antwortverzerrungen

Unter Antwortverzerrung versteht man die systematische Tendenz einer Person, in Testverfahren Antworten zu geben, die nicht mit der tatsächlichen Merkmalsausprägung übereinstimmen (Herzberg, 2011). Es werden mehrere Arten von Antwortverzerrungen unterschieden, wobei die Begriffe in der Literatur nicht immer einheitlich verwendet werden.

Unter negativen Antwortverzerrungen (Fake-bad) wird die Neigung verstanden, vorhandene Auffälligkeiten und Symptome zu übertreiben (Aggravation) oder das Vorhandensein von Beschwerden und Störungen vorzutäuschen (Simulation) sowie das Präsentieren von zu schlechten Leistungen (suboptimales Leistungsverhalten; Merten et al., 2009). Zu den positiven Antwortverzerrungen (Fake-good) gehören nach Rogers (2008a) und Herzberg (2011) zum einen der Versuch, vorhandene Symptome oder Störungen zu untertreiben oder verleugnen (Dissimulation), zum anderen die Neigung, sich selbst in einem entsprechend sozialer Normen und Werte möglichst günstigen Licht zu präsentieren (soziale Erwünschtheit). Das Konstrukt der sozialen Erwünschtheit wird von Paulhus (2002) weiter unterteilt in Selbsttäuschung (Self-deception), eine positiv verzerrte, aber subjektiv als ehrlich empfundene Selbstdarstellung, und Impression Management, die Steuerung der Wahrnehmung der eigenen Person durch das Gegenüber durch bewusste Antwortverfälschung. Von Cima und Kollegen (Cima, Hollnack, Kremer, Knauer & Merckelbach, 2003a) wurde für den forensischen Begutachtungskontext der Begriff Supernormalität eingeführt. Zunächst wurde sie als das „absichtliche zur Schau stellen oder Übertreiben positiver Verhaltensweisen und sozial erwünschter Meinungen“ (S. 113) definiert und explizit von Dissimulation und Sozialer Erwünschtheit abgegrenzt. Die Abgrenzung wurde damit begründet, dass im Gegensatz zu Dissimulation „aktives, bewusstes Vortäuschen normaler und gesunder Verhaltensweisen“ (Cima et al., 2003a, S. 114) bzw. nicht nur die Verleugnung psychiatrischer

Symptome, sondern die Verleugnung verbreiteter Beschwerden erfasst wird (Cima et al., 2008). Soziale Erwünschtheit unterscheide sich von Supernormalität insofern, als erstere sich allgemein auf soziales Verhalten beziehe, „ohne dass unbedingt eine psychische Störung vorliegt“ (Cima et al., 2003a, S. 114), bzw. als Supernormalität nicht nur vom sozialen Kontext abhängig sei (Cima et al., 2008). Diese Abgrenzung wurde in neueren Publikationen nicht mehr vorgenommen, sondern Dissimulation wurde synonym verwendet und soziale Erwünschtheit als eng verbundenes Merkmal definiert (Cima, Pantus & Dams, 2009). Das Konstrukt wurde dann definiert als das systematische Leugnen und Abstreiten alltäglicher und weit verbreiteter Beschwerden, eine Beschwerdenminimierung sowie das absichtliche Vortäuschen und Übertreiben positiver Eigenschaften bzw. gesunder Merkmale.

Die oben genannten Antwortverzerrungen sind abhängig vom Inhalt der bearbeiteten Items. Es gibt aber auch Antwortverzerrungen, die unabhängig davon sind und sich auf die Unfähigkeit oder den fehlenden Willen des Probanden beziehen, die Items zu verstehen oder sich damit auseinanderzusetzen (Sellbom & Bagby, 2008). Dazu gehören z.B. die Ja-Sage- oder Zustimmungstendenz (Akquieszenz) und die Nein-Sage-Tendenz, also das Zustimmung zu oder Ablehnen von Items unabhängig vom Inhalt, sowie das unsystematische oder zufällige Antworten. Personen mit Tendenz zur Mitte wählen unabhängig vom Iteminhalt überwiegend die mittlere Antwortkategorie.

Von einigen Autoren wurden die in einer spezifischen Situation oder bei einem bestimmten Testverfahren auftretenden Verzerrungen, sog. Response Sets, von den über verschiedene Verfahren und Untersuchungskontexte hinweg auftretenden und von spezifischen Stimuli eher unabhängigen Response Styles unterschieden (Paulhus, 1991). Amelang und Schmidt-Atzert (2006) kritisieren den allzu inkonsistenten Gebrauch dieser zwei Begriffe und verwenden den Terminus Antwortstil für jegliche Tendenz einer Person, „Items aus einem Fragebogen nicht (nur) nach dem Wahrheits-, sondern nach einem anderen Prinzip zu beantworten, das nicht der Intention der Testautoren entspricht“ (S. 247). Auch im angloamerikanischen Raum wird mittlerweile häufig der Begriff Response Styles als Oberbegriff für die o.g. Verzerrungsarten verwendet. Denn nach Rogers (2008a) ist die Entscheidung einer Person, ob sie sich in einer Untersuchungssituation öffnet und ehrlich antwortet oder aber den Untersucher täuscht, häufig eine rationale, die von ganz unterschiedlichen Einflussfaktoren determiniert wird, z.B. durch die Beziehung zum Untersucher, individuelle Ziele der Person oder Merkmale der Situation. Antwortverzerrungen sind demnach kein stabiles Persönlichkeitsmerkmal einer bestimmten Person (trait). Damit kann eine Person in dersel-

ben Untersuchungssituation verschiedene Antwortverzerrungen einsetzen (Lanyon & Cunningham, 2005; Otto, 2008). Dies wird als hybrides Antwortverhalten (Rogers, 2008a) bezeichnet.

Da in der vorliegenden Arbeit vor allem das bewusste Täuschen des Untersuchers bzw. die bewusste Verfälschung testpsychologischer Verfahren bei forensischen Begutachtungen untersucht wurde, wird hier die im Vergleich zu Response Style weniger neutrale Bezeichnung Antwortverzerrung verwendet.

2.2 Positive Antwortverzerrungen im forensischen Begutachtungskontext

Das Thema Antwortverzerrungen bei Begutachtungen wurde im deutschsprachigen Raum lange Zeit eher tabuisiert (Merten et al., 2009). Seit den 1990er Jahren widmeten jedoch in der Forschung und auch in der Praxis tätige forensische Psychologen und Psychiater diesem Bereich zunehmend Aufmerksamkeit. Allerdings lag der Schwerpunkt dabei eindeutig auf den negativen und weniger auf den positiven Antwortverzerrungen (Otto, 2008). Empirische Erkenntnisse zu positiven Antwortverzerrungen, insbesondere zur sozialen Erwünschtheit, entstammen vorwiegend dem Bereich der Personalauswahl (Borkenau, Friedel & Wolfradt, 2011; Griffith & Peterson, 2006), aus der forensischen Psychologie und Psychiatrie liegen eher wenige Studien vor (Dettenborn, 2009; Otto, 2008).

2.2.1 Häufigkeit von positiven Antwortverzerrungen bei forensischen Begutachtungen

Während für die Prävalenz negativer Antwortverzerrungen bei Begutachtungen verschiedene Schätzungen und empirische Studien vorliegen (Übersicht bei Merten, 2014; Merten et al., 2009), ist die Auftretenshäufigkeit positiver Antwortverzerrungen bislang kaum untersucht. Es ist jedoch anzunehmen, dass die Prävalenz nicht unerheblich und in bestimmten Kontexten sogar höher ist als die negativer Antwortverzerrungen (Rogers, 2008b).

Grossman und Wasyliw (1988) untersuchten die Ergebnisse im Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1951) von 49 Probanden, die im Rahmen des Strafverfahrens zur Frage der psychischen Gesundheit zum Tatzeitpunkt und/oder der Verhandlungsfähigkeit begutachtet wurden. Je nach Validitätsskala, die zur Beurteilung des Antwortverhaltens herangezogen wurde, zeigten zwischen 14% und 41% negative Antwortverzerrungen und zwischen 8% und 22% positive Antwortverzerrungen. Heilbrun, Bennett, White und Kelly (1990) untersuchten das Antwortverhalten von 159 männlichen psychisch kranken Rechtsbrechern, die in einer forensischen Klinik unterge-

bracht waren, mit dem MMPI, der von den Patienten innerhalb von drei bis fünf Wochen nach Aufnahme bearbeitet worden war. Nur bei 22% wurden die Selbsteinschätzungen als zuverlässig klassifiziert, 46% der Männer zeigten Antwortverzerrungen. Hier traten positive Antwortverzerrungen deutlich häufiger auf (21%) als negative (9%).

Baer und Miller (2002) schätzen in ihrer Übersichtsarbeit zu positiven Antwortverzerrungen im MMPI-2 (Hathaway & McKinley, 1989) die Basisrate auf 30%, sowohl für den Bereich der Personalauswahl als auch für Begutachtungen im Rahmen eines Sorgerechtsstreits.

Aufbauend auf der Annahme, dass es sich bei Antwortverzerrungen nicht um stabile Persönlichkeitsmerkmale handelt, sondern um ein situationsabhängiges Verhalten, ist davon auszugehen, dass die Prävalenz von verschiedenen Einflussfaktoren abhängt.

2.2.2 Einflussfaktoren auf das Antwortverhalten

Für den forensischen Bereich wurden Unterschiede im Antwortverhalten insbesondere in Abhängigkeit vom Untersuchungskontext und des Vorliegens einer Psychopathie untersucht. Im Folgenden sollen vor allem die empirischen Befunde zu positiven Antwortverzerrungen bei forensischen Probanden vorgestellt werden. Es werden aber auch relevante Ergebnisse aus anderen Bereichen referiert, vor allem solche, welche die Wahl der Methoden dieser Arbeit beeinflussten. Die Arbeit beschränkt sich auf die Darstellung von Antwortverhalten im Bereich der Persönlichkeitsdiagnostik. Denn positive Antwortverzerrungen bzw. Täuschungsversuche bei der Diagnostik von intellektueller und kognitiver Leistungsfähigkeit im Sinne eines Verbergens von Defiziten, etwa durch gezielte Vorbereitung oder Coaching, spielen im forensischen Bereich eine untergeordnete Rolle (Rogers, 2008b).

2.2.2.1 *Untersuchungskontext bzw. -zweck*

Ob eine Person positive Antwortverzerrungen zeigt, ist abhängig von verschiedenen Merkmalen des Untersuchungskontextes bzw. der Untersuchungsbedingungen.

In der empirischen Sozialforschung wird zunehmend auf Online-Befragungen zurückgegriffen, da davon ausgegangen wird, dass Probanden in anonymen Testsituationen, d.h. wenn die Ergebnisse der Untersuchung nicht ihrer Person zugeordnet werden können, weniger sozial erwünscht antworten (vgl. Maurer & Jandura, 2009; Kaufmann & Reips, 2008; Taddicken, 2009). So fand beispielsweise Paulhus (1984), dass Impression Management, nicht aber Self-Deception, in einer anonymen Gruppenuntersuchung seltener auftritt, als

bei Versuchspersonen, die ihre Personalien angeben und den ausgefüllten Fragebogen direkt beim Untersucher abgeben müssen. In einer Metaanalyse konnten Weisband und Kiesler (1996) zeigen, dass die Selbstöffnungsbereitschaft von Probanden in Onlinebefragungen, die über ein höheres Ausmaß an wahrgenommener Anonymität verfügen (vgl. Taddicken, 2009), höher ist als in Face-to-Face-Untersuchungen; insbesondere dann, wenn sensible Daten erhoben werden (z.B. zur psychischen Gesundheit oder Kriminalität) oder Populationen untersucht wurden, für die die Testergebnisse besondere Konsequenzen haben konnten (z.B. Psychiatriepatienten).

Allerdings hat der Grad der Anonymisierung von Befragungen lediglich für den Forschungskontext (s. Kap. 2.4.2), nicht aber für forensische Begutachtungen praktische Relevanz. Hier beeinflussen andere Merkmale des Untersuchungskontextes das Ausmaß an positiven Antwortverzerrungen.

Der Kontext bzw. die Fragestellung der Untersuchung beeinflusst die Auftretenswahrscheinlichkeit insofern, als Ergebnisse einer Untersuchung für Probanden eine unterschiedliche Bedeutsamkeit haben können, was die Motivation zu einer verzerrten Selbstpräsentation beeinflusst. Je bedeutsamer die Testergebnisse für eine Person sind, desto höher dürfte der Anreiz, die Befunde in die gewünschte Richtung zu beeinflussen, bzw. die Verfälschungsentention sein (Herzberg, 2011; Mueller-Hanson, Heggstad & Thornton III, 2006). Da die Befunde von Begutachtungen eine hohe Bedeutsamkeit für die Probanden haben, ist demnach auch mit einer höheren Bereitschaft zu Antwortverzerrungen zu rechnen.

Innerhalb des forensischen Settings variiert die Auftretenshäufigkeit und Art von Antwortverzerrungen mit der Fragestellung der Begutachtung (Rogers, 2008a). Je nachdem, welches Ziel Probanden bei der Begutachtung verfolgen, werden Beschwerden, Symptome, Auffälligkeiten etc. eher besonders hervorgehoben oder aber heruntergespielt oder verleugnet (Greene, 2007). Wenn Probanden in einem sozial- oder zivilrechtlichen Verfahren auf eine Zahlung wegen psychischer Beeinträchtigungen oder Störungen hoffen, erhöhen sie die Wahrscheinlichkeit, ihr Ziel zu erreichen, wenn sie Beschwerden herausstellen oder übertreiben (increased reporting of symptoms). Dies gilt auch für strafrechtliche Verfahren, wenn Probanden versuchen, wegen verminderter Schuldfähigkeit aufgrund einer psychischen Störung eine Strafmilderung zu erhalten oder einer Gefängnisstrafe durch Unterbringung in einem psychiatrischen Krankenhaus zu entgehen. Demgegenüber besteht in Kontexten, in denen Probanden bestrebt sind, psychische Gesundheit nachzuweisen, eine höhere Wahrscheinlichkeit für das Auftreten von Minimierungs- und Leugnungstendenzen (decreased reporting of symptoms). Der Anreiz zu positiven Antwortverzerrungen ist also z.B.

besonders hoch, wenn der Entzug des Sorgerechts droht, gegen den Willen von Probanden eine gesetzliche Betreuung eingerichtet werden soll oder Probanden vorzeitig bzw. auf Bewährung aus der Haft oder dem Maßregelvollzug entlassen werden möchten. Dass Probanden zielgerichtet ihr Antwortverhalten je nach Untersuchungskontext anpassen können, illustriert Greene (2007) am Beispiel der MMPI-Profile eines Probanden, der zweimal im Rahmen eines Zivilprozesses als Opfer einer Körperverletzung und zwischen diesen beiden Begutachtungen im Rahmen eines Sorgerechtsstreits untersucht wurde. Während er im Kontext des Verfahrens wegen Körperverletzung jeweils eine Vielzahl von körperlichen und psychischen Symptomen berichtete, gab er im Rahmen des Sorgerechtsverfahrens nur wenige körperliche und kaum psychische Symptome an, was auf einzelnen Skalen zu deutlich unterschiedlichen Ergebnissen mit einer Differenz von mehr als drei Standardabweichungen führte. Walters (1988) konnte anhand der MMPI-Validitätsskalen zeigen, dass Insassen eines Hochsicherheitsgefängnisses je nach Untersuchungszweck unterschiedliche Antwortverzerrungen zeigten. Häftlinge, die den Antrag auf Unterbringung in einer Einzelzelle gestellt hatten (Fake-bad-Bedingung), zeigten häufiger negative Antwortverzerrungen, während Häftlinge, die zur Frage der bedingten Haftentlassungen untersucht wurden (Fake-good-Bedingung), häufiger positive Antwortverzerrungen zeigten. Häftlinge, die im Rahmen der routinemäßigen Eingangsdiagnostik für eine Gruppentherapie untersucht wurden (Kontrollgruppe), unterschieden sich hinsichtlich des Antwortverhaltens deutlich von den Probanden der Fake-bad-Bedingung, neigten insgesamt aber auch häufig zu Fake-good-Verhalten.

Das Ausmaß der Verfälschungsintention einer Person ist auch davon abhängig, ob sie im konkreten Untersuchungskontext eine Verfälschung in der gewünschten Richtung überhaupt als möglich ansieht bzw. tatsächlich in der Lage dazu ist (Mueller-Hanson et al., 2006). Dies wiederum ist sowohl von Merkmalen der Probanden selbst als auch von den in der Untersuchung eingesetzten diagnostischen Verfahren abhängig. Auf beides wird in den folgenden beiden Abschnitten eingegangen.

2.2.2.2 *Probandenmerkmale*

Das Auftreten von Antwortverzerrungen ist nicht nur abhängig vom Untersuchungskontext, sondern auch von bestimmten Merkmalen der untersuchten Personen. Von einigen Autoren wurde Faking good bzw. soziale Erwünschtheit, wie bereits erwähnt, sogar als spezifisches Persönlichkeitsmerkmal konzeptualisiert (Übersicht bei Herzberg, 2011). Dies wurde u.a. damit begründet, dass sich die Übereinstimmung zwischen Selbst- und Fremd-

beurteilungen von bestimmten Persönlichkeitsmerkmalen durch Auspartialisieren von Kontrollskalen nicht verbessert (McCrae & Costa, 1983; Pauls & Stemmler, 2003) und soziale Erwünschtheit auch ohne Anreizbedingungen substantiell mit emotionaler Stabilität und Gewissenhaftigkeit korreliert (Nicholson & Hogan, 1990; Ones, Viswesvaran & Reiss, 1996).

Mueller-Hanson et al. (2006) kritisierten, dass zwar bekannt ist, dass es individuelle Unterschiede hinsichtlich des Auftretens von Täuschungsverhalten gibt, aber bisher kaum empirisch fundierte Modelle, die diese zufriedenstellend erklären. Sie entwickelten aufbauend auf früheren Modellen und bisherigen empirischen Befunden ein integratives Modell zur Entstehung von positiven Antwortverzerrungen. Darin wird angenommen, dass (a) die Wahrnehmung der Situation (subjektive Bedeutung des Täuschungsverhaltens, wahrgenommene Fähigkeit, erfolgreich zu täuschen, subjektive Normen/Annahmen über die Angemessenheit von Täuschungsverhalten), (b) die Bereitschaft zu täuschen, (c) die Fähigkeit zu täuschen sowie (d) die Persönlichkeitsmerkmale Gewissenhaftigkeit und emotionale Stabilität die konkrete Täuschungsabsicht beeinflussen, welche das tatsächliche Täuschungsverhalten bedingt. Die Autoren gehen davon aus, dass sowohl die Fähigkeit als auch die Bereitschaft einer Person zu täuschen von spezifischen Persönlichkeitsmerkmalen beeinflusst wird. Anhand einer studentischen Stichprobe in einer simulierten Einstellungssituation konnten sie u.a. zeigen, dass die Persönlichkeitsmerkmale Machiavellismus, geringes Regelbewusstsein und gutes Self-Monitoring Dispositionen für die Bereitschaft zu Faking-good-Verhalten darstellen. Allerdings war in ihrer Studie dieses zwar mit der Situationswahrnehmung, nicht aber mit der von den Probanden berichteten Täuschungsabsicht assoziiert.

In Übereinstimmung mit dem Modell von Mueller-Hanson et al. (2006) wurde im forensischen Bereich vor allem der Zusammenhang zwischen psychopathischen Persönlichkeitsmerkmalen und Antwortverzerrungen näher untersucht. Hintergrund war, dass pathologisches Lügen und betrügerisch-manipulatives Verhalten zu den Charakteristika der Psychopathie gehören (Hare, 1991) und daher eine höhere Wahrscheinlichkeit für das Auftreten positiver und negativer Antwortverzerrungen besteht, und dass Personen mit hohen Psychopathiewerten erfolgreicher täuschen als Personen mit niedrigen Werten. Die empirischen Befunde sind jedoch keineswegs konsistent. So konnten z.B. Gacono, Meloy, Sheppard, Speth und Roske (1995) zeigen, dass erfolgreiche Simulanten, die aufgrund einer vorgetäuschten psychischen Störung schuldunfähig gesprochen und in einem psychiatrischen Krankenhaus untergebracht worden waren, häufiger eine Antisoziale Persönlich-

keitsstörung bzw. höhere Werte in der Psychopathy Checklist-Revised (PCL-R; Hare, 1991) hatten. Dagegen fanden Poythress, Edens und Watkins (2001) in einer Stichprobe von Strafgefangenen keinen Zusammenhang zwischen Psychopathie, erfasst mit dem Psychopathic Personality Inventory (PPI; Lilienfeld & Andrews, 1996), und Simulation psychischer Störungen. Kucharski, Duncan, Egan und Falkenbach (2006) berichteten, dass in einer Stichprobe angeklagter, aber noch nicht verurteilter Täter, die im Auftrag des Gerichts psychologisch begutachtet wurden, Probanden mit hohen PCL-R-Werten höhere Werte auf verschiedenen Skalen zu negativen Antwortverzerrungen erzielten als Probanden mit niedrigen PCL-R-Werten. Allerdings war eine zuverlässige Klassifikation von Simulanten und Nicht-Simulanten aufgrund der Diagnose Psychopathie nicht möglich, da die meisten Probanden selbst mit sehr hohen PCL-R-Werten keine negativen Antwortverzerrungen zeigten.

Die Zusammenhänge zwischen positiven Antwortverzerrungen und Psychopathie wurden bislang kaum und vorwiegend an studentischen Stichproben untersucht (mit Ausnahme der Studien zur Supernormalität-Skala, s. Kap. 2.5). Edens, Buffington, Tomicic und Riley (2001) untersuchten anhand einer Stichprobe nicht-krimineller Erwachsener die Auswirkungen von positiven Antwortverzerrungen auf die PPI-Werte, indem sie die PPI-Werte unter einer Standardinstruktion mit den PPI-Werten unter einer Fake-Instruktion verglichen. Bei den Probanden mit hohen Psychopathiewerten zeigte sich eine größere Diskrepanz der PPI-Werte zwischen den beiden Instruktionsbedingungen, was darauf hindeuten könnte, dass psychopathische Personen erfolgreicher dissimulieren können. Dagegen fanden MacNeil und Holden (2006) in einer studentischen Stichprobe keinen Zusammenhang zwischen den selbstberichteten Psychopathiewerten und erfolgreichem Täuschungsverhalten. Die Personen, die auch unter Fake-good-Instruktion auf den Kontrollskalen unauffällige Werte erzielten (successful fakers), hatten lediglich auf den PPI-Skalen Machiavellistischer Egoismus und Schuldexternalisierung signifikant höhere Werte als Personen, deren Antwortverhalten durch die Kontrollskalen als sozial erwünscht eingestuft wurde (unsuccessful fakers).

In einer Metaanalyse von 45 Studien fanden Ray, Hall, Rivera-Hudson, Poythress, Lilienfeld und Morano (2013) negative Zusammenhänge zwischen selbstberichteten psychopathischen Merkmalen und positiven Antwortverzerrungen sowie positive Zusammenhänge mit negativen Antwortverzerrungen. Es waren aber ausschließlich Studien eingeschlossen, in denen die Ergebnisse, im Gegensatz etwa zu strafrechtlichen Begutachtungen, keine persönliche Bedeutung für die Probanden hatten, also kein spezifischer Anreiz für Ant-

wortverzerrungen bestand, was die Generalisierbarkeit der Ergebnisse auf die Begutachtungspraxis fraglich erscheinen lässt. Außerdem wurden Studien mit Fremdratings nicht eingeschlossen, sodass die Befunde auch damit erklärt werden könnten, dass von den untersuchten Probanden aufgrund positiver Antwortverzerrungen niedrige Psychopathiewerte und aufgrund negativer Antwortverzerrungen hohe Psychopathiewerte erzielt wurden. Edens und Kollegen (Edens, Buffington & Tomicic, 2000; Edens et al., 2001) konnten nämlich anhand studentischer Stichproben zeigen, dass die selbstberichteten Psychopathiewerte unter einer Fake-good-Instruktion sinken, während sie unter einer Fake-bad-Instruktion steigen.

Neben den methodischen Einschränkungen der bislang durchgeführten Studien zum Zusammenhang zwischen Psychopathie und Fake-good ist auch zu berücksichtigen, dass bislang der Einfluss einer unbewussten Selbsttäuschung (Paulhus, 2002), der Introspektionsfähigkeit bzw. der verzerrten Selbstwahrnehmung nicht untersucht wurde (Steinmayr & Amelang, 2011). Dies ist insbesondere bei der Untersuchung von Zusammenhängen zwischen Persönlichkeitspathologie und Antwortverzerrungen relevant, da eine verzerrte Selbstwahrnehmung bzw. ein Mangel an Einsicht in das eigene Problemverhalten Kernmerkmal von Persönlichkeitsstörungen ist. Niedrige Werte auf Skalen zu problematischen Persönlichkeitseigenschaften bzw. die geringe Übereinstimmung zwischen Selbst- und Fremdbeurteilung bei Probanden mit Persönlichkeitsstörung (Fiduccia, 2011; Whyte, Fox & Coxell, 2006) könnten demnach nicht nur auf bewusste positive Antwortverzerrungen, sondern auch auf eine mangelnde Krankheitseinsicht zurückzuführen sein. In solchen Fällen ist jedoch nicht nur mit niedrigen Werten auf Inhaltsskalen, sondern auch mit auffälligen Werten auf Kontrollskalen zu rechnen, wenn Inhalte erfasst werden, für welche die Probanden aufgrund ihrer Störung keine validen Selbstangaben liefern können (Steinmayr & Amelang, 2011).

Für negative Antwortverzerrungen konnte bereits gezeigt werden, dass Validitätsskalen oder Beschwerdenvalidierungsverfahren nicht nur bewusste Verfälschungsversuche, sondern z.T. eher das Ausmaß der Psychopathologie erfassen (Wetzler & Marlowe, 1990). Personen mit psychischen Störungen erzielen in Untersuchungssituationen ohne Anreizbedingungen in Skalen zu negativen Antwortverzerrungen höhere Werte als gesunde Kontrollprobanden (Veress, 2010; Mader, 2014), was dafür spricht, dass solche Verfahren nicht nur Täuschungsversuche, sondern auch eine tatsächlich vorhandene Symptombelastung messen.

Grillo (1994) analysierte die Zusammenhänge zwischen den Ergebnissen eines Selbstbeurteilungsfragebogens zu Persönlichkeitsstörungen (Millon Clinical Multiaxial Inventory-II [MCMI-II]; Millon, 1992) und verschiedenen Validitätsskalen des MMPI-2 anhand einer Stichprobe von Begutachtungsprobanden mit Schadensersatzforderungen. Mit Ausnahme der Skalen zur dependenten und narzisstischen Persönlichkeitsstörung korrelierten alle MCMI-Skalen signifikant mit Validitätsskalen des MMPI. Der Autor schließt daraus, dass die erhöhten Werte forensischer Probanden also nicht nur auf negative Antwortverzerrungen, sondern auch auf die spezifische Psychopathologie zurückzuführen sind.

Unter der Annahme, dass bestimmte Persönlichkeitsstile mit bestimmten Antwortstilen assoziiert sind bzw. dass ein bestimmtes Antwortverhalten im Rahmen einer spezifischen Persönlichkeitsstörung typischerweise auftritt, d.h. Teil des Störungsbildes und nicht Ausdruck eines bewussten Täuschungsversuches ist, untersuchte Wise (2002) eine Stichprobe von inhaftierten Angeklagten, die im Auftrag des Gerichts psychologisch untersucht wurden, ebenfalls mit MCMI und MMPI, wobei Probanden mit ungünstigem Profil ausgeschlossen wurden. Entsprechend der Hypothese korrelierte jede der Validitätsskalen signifikant mit mindestens zwei Skalen zu Persönlichkeitsstörungen. Dabei waren Persönlichkeitsstörungen des Clusters B (dramatisch, launisch, emotional) des DSM-IV (Saß, Wittchen, Zaudig & Houben, 2003) vor allem mit negativen Antwortverzerrungen assoziiert, während sich zwischen Cluster C-Persönlichkeitsstörungen (ängstlich, vermeidend) und Antwortverzerrungen kaum bzw. gar keine Zusammenhänge zeigten. Die meisten und stärksten Zusammenhänge wiesen aber alle Persönlichkeitsstörungen mit Faking-bad-Skalen auf. Systematische Zusammenhänge zwischen einzelnen Persönlichkeitsstörungen und spezifischen Antwortverzerrungen konnten nicht gefunden werden.

Ein weiterer Faktor, der beeinflusst, ob eine Person Antwortverzerrungen zeigt, ist ihre Fähigkeit, bewusst zu täuschen. Die Person muss erkennen können, in welchen Verfahren bzw. auf welchen Items sie welche Antworten geben muss, um die erwünschten Testergebnisse zu erzielen bzw. um sich erfolgreich, so wie geplant, zu präsentieren. Häufig wurde angenommen, dass die kognitiven Fähigkeiten bzw. die Intelligenz daher eine große Rolle bei der Entstehung von Antwortverzerrungen spielen (Herzberg, 2011). Die empirischen Befunde sind aber uneinheitlich (Mueller-Hanson et al., 2006). Ones et al. (1996) fanden in einer Metaanalyse keinen Zusammenhang zwischen den kognitiven Fähigkeiten und sozialer Erwünschtheit. Die Fähigkeit, Ergebnisse in der gewünschten Richtung zu beeinflussen, dürfte also weniger von der Intelligenz im Allgemeinen als vielmehr von dem Wissen abhängig sein, wie in dem spezifischen Untersuchungskontext in dem einge-

setzten Verfahren geantwortet werden muss, um beim Untersucher den gewünschten Eindruck zu erzielen (Mueller-Hanson et al., 2006). Somit hat die Art des eingesetzten Untersuchungsmaterials entscheidenden Einfluss auf die Auftretenswahrscheinlichkeit.

2.2.2.3 *Untersuchungsmaterial*

Grundsätzlich sind alle (testpsychologischen) Verfahren, die auf Auskünften von Probanden beruhen, anfällig für Antwortverzerrungen. Sie können umso leichter verfälscht werden, je leichter das Messprinzip bzw. die Messintention für die Testperson zu durchschauen ist (Kubinger, 2006) bzw. je höher die Augenscheinvalidität ist (Bornstein, Rossner, Hill & Stepanian, 1994). Dementsprechend wird traditionell angenommen, dass projektive Verfahren, wie der Thematische Apperzeptionstest (TAT; Murray, 1991; Revers, 1993) oder das Rorschach-Verfahren (Bohm, 2004; Rorschach, 1994), die unstrukturiertes Stimulusmaterial nutzen, weniger leicht zu durchschauen und damit weniger anfällig für Verfälschungstendenzen sind als Fragebögen (Sewel, 2008; Weiner, 2007). Deshalb kommt ihnen im forensischen Bereich nach wie vor Bedeutung zu (Scheurer & Richter, 2005), auch wenn ihr Einsatz in Wissenschaft und (klinischer) Praxis kontrovers diskutiert wird (Lilienfeld, Wood & Garb, 2000; Wittkowski, 2011).

Empirische Studien, in denen verschiedene Verfahren mit demselben Messanspruch hinsichtlich der Verfälschbarkeit miteinander verglichen wurden, sind jedoch selten, nicht zuletzt da nur für wenige Konstrukte verschiedene validierte Methoden verfügbar sind. Ziegler, Schmidt-Atzert, Bühner und Krumm (2007) konnten an einer studentischen Stichprobe zeigen, dass sich Selbstbeurteilungsverfahren, projektive Verfahren und objektive Verfahren (s.u.) zur Erfassung von Leistungsmotivation hinsichtlich der Verfälschbarkeit unterscheiden. Während das eingesetzte objektive Verfahren durch positive Antwortverzerrungen nicht zu verfälschen war, unterschieden sich die beiden anderen Methoden hinsichtlich des Ausmaßes an Verfälschbarkeit: Die Ergebnisse der Selbstbeurteilungsskala wurden durch Fake-good-Instruktionen stärker verzerrt als die des projektiven Verfahrens. Bornstein et al. (1994) untersuchten anhand einer studentischen Stichprobe den Zusammenhang zwischen Augenscheinvalidität und Verfälschbarkeit von Verfahren zur Erfassung des Persönlichkeitsmerkmals Abhängigkeit. Das eingesetzte Selbstbeurteilungsinventar hatte eine höhere Augenscheinvalidität und war durch Fake-Instruktionen leichter zu verfälschen als der Abhängigkeits-Index des Rorschach-Verfahrens, der eine geringere Augenscheinvalidität hatte. Biedermann (2007) untersuchte eine Stichprobe inhaftierter jugendlicher Gewalttäter mit Selbstbeurteilungs- und projektiven Verfahren zu Aggressivi-

tät. Eine Fremdbeurteilung der Aggressivität durch Betreuungspersonen im Justizvollzug (Außenkriterium) wies häufiger signifikante Beziehungen zu den Ergebnissen der projektiven Verfahren als zu den Ergebnissen von Selbstbeurteilungsskalen auf.

Täuschungsversuche scheinen zwar einen geringeren Einfluss auf die Befunde projektiver Verfahren als auf die Befunde von Selbstbeurteilungsverfahren zu haben. Aber es gibt nicht nur kaum Studien, in denen der Einfluss von positiven Antwortverzerrungen auf die Befunde bzw. die klinische Interpretation der Ergebnisse projektiver Verfahren untersucht wurde (Übersicht bei Sewell, 2008), sondern es wurden bislang auch keine zuverlässigen Methoden entwickelt, um Antwortverzerrungen in häufig eingesetzten projektiven Verfahren zuverlässig zu identifizieren. Nicht zuletzt deshalb werden für forensische Fragestellungen immer mehr objektive Persönlichkeitstests bzw. indirekte Verfahren entwickelt. Bei solchen Verfahren werden Persönlichkeitsmerkmale aus dem beobachtbaren Verhalten einer Person in einer standardisierten Situation erschlossen (Kubinger, 2006), weshalb sie als wenig bis kaum zu verfälschen gelten. Gerade im Bereich der Diagnostik von sexuellen Präferenzen und Aggression sind in den letzten Jahren eine Reihe von indirekten Verfahren entwickelt und validiert worden (Briken, Rettenberger & Dekker, 2013; Smith & Waterman, 2004; Werner, 2005; Yundina & Nedopil, 2010).

Dass insbesondere im forensischen Bereich großes Interesse an Verfahren mit geringer Durchschau- und Verfälschbarkeit herrscht, ist nachvollziehbar angesichts der Tatsache, dass auch der Testinhalt bzw. Messgegenstand das Ausmaß von Antwortverzerrungen beeinflusst. Antwortverzerrungen treten nicht skalenunabhängig auf. Vor allem bei der Erfassung stark negativ konnotierter Konstrukte, d.h. solcher Eigenschaften, die aufgrund soziokultureller Normen als „schlecht“ oder gar „böse“ gelten, z.B. Machiavellismus, Narzissmus, Psychopathie (Steinmayr & Amelang, 2011), ist stärker oder häufiger mit dem Auftreten positiver Antwortverzerrungen zu rechnen. Gerade solch „sozial unerwünschte“ bzw. von der Norm deutlich abweichende Persönlichkeitsmerkmale oder Verhaltensdispositionen werden aber bei strafrechtlichen Begutachtungen erfasst, da sie mit Delinquenz assoziiert sind bzw. kriminogene Merkmale darstellen (Andrews & Bonta, 2010). Um die Wahrscheinlichkeit bzw. das Ausmaß von Antwortverzerrungen zu reduzieren, sollten daher – entsprechend der Empfehlungen von Hemphill und Hart (2003) zur Diagnostik von Psychopathie – für forensische Begutachtungen Selbstbeurteilungsverfahren mit möglichst neutralen oder gar positiven Itemformulierungen ausgewählt werden.

2.2.3 Auswirkung von Antwortverzerrungen auf testpsychologische Befunde

Antwortverzerrungen haben einen Einfluss auf psychometrische Eigenschaften von Verfahren. So kann soziale Erwünschtheit die faktorielle Struktur von Persönlichkeitsverfahren verändern, die Reliabilität bzw. interne Konsistenz sowie die kriteriumsbezogene Validität mindern (Übersicht bei Herzberg, 2011). Auswirkungen auf die Ergebnisse von Selbstbeurteilungsverfahren wurden in den vorangehenden Kapiteln z.T. bereits berichtet. Bei forensischen Begutachtungen zeigt sich immer wieder, dass die Skalenwerte von Probanden mit positiven Antwortverzerrungen auffallend niedrig (bei „unerwünschten“ Eigenschaften) oder hoch (bei „erwünschten“ Eigenschaften) ausgeprägt sind und die Befunde nicht mit der Fremdeinschätzung durch den Untersucher selbst oder Dritte, wie die behandelnden Therapeuten, übereinstimmen (Tippelt et al., 2010). In einer Reihe von Studien, von denen hier einige exemplarisch dargestellt werden, wurden solche Auswirkungen von Antwortverzerrungen auf testpsychologische Ergebnisse näher untersucht. Allerdings handelt es sich dabei überwiegend um experimentelle Untersuchungen. In einer entsprechenden Metaanalyse zur Verfälschbarkeit der Big Five Persönlichkeitsfaktoren unterschieden sich die Skalenwerte unter Fake-Instruktion im Vergleich zur Standardinstruktion um rund eine halbe Standardabweichung (Viswesvaran & Ones, 1999). Wie bereits erwähnt, konnten Edens und Kollegen (Edens, 2004; Edens et al., 2001) zeigen, dass Probanden nach einer Fake-Instruktion signifikant niedrigere Gesamtwerte im PPI erzielen als mit der Standardinstruktion. Mit klinischen oder forensischen Probanden wurden weniger Studien durchgeführt. Die Auswirkungen von Fake-good bei forensischen Probanden werden häufig durch den Vergleich von Testergebnissen zwischen Personen mit auffälligen und unauffälligen Werten auf Skalen zu Antwortverzerrungen untersucht, aber auch mit Fake-Instruktion. Grossman, Wasyliv, Benn und Gyoerkoe (2002) analysierten die MMPI-Profile von Männern, die eines Sexualdelikts beschuldigt und zu einer forensisch-psychologischen Begutachtung überwiesen worden waren. Probanden mit auffälligen Werten auf MMPI-Dissimulationsskalen („minimizers“) erzielten auf sechs (MMPI) bzw. fünf (MMPI-2) der zehn klinischen Skalen signifikant niedrigere Werte als Probanden mit unauffälligen Werten auf den entsprechenden Validitätsskalen. An einer Stichprobe von forensischen Probanden, die wegen Stalkingverhaltens begutachtet wurden, zeigten McEwan, Davis, MacKenzie und Mullen (2009), dass Probanden mit hohen Werten auf einer Impression Management-Skala signifikant weniger dispositionellen Ärger und mehr Ärgerkontrolle berichteten als Probanden mit niedrigen IM-Werten. Steffan, Kroner und Morgan (2007) untersuchten Haftinsassen mit einem Persönlichkeitsinventar zur Erfassung

von Psychopathologie zu Beginn der Inhaftierung im Rahmen der Eingangsdiagnostik mit Standardinstruktion und einen Monat später mit Fake-good-Instruktion. Auf allen neun klinischen Skalen wurden in der Fake-Bedingung niedrigere Werte erzielt, die Effektstärken waren jedoch mit Ausnahme der Skalen Interpersonelle Probleme und Entfremdung eher gering, was die Autoren damit begründen, dass die Probanden bereits in der Standard-Bedingung versuchten, sich in möglichst vorteilhaftem Licht zu präsentieren, aber auch damit, dass positive Antwortverzerrungen mit testpsychologischen Verfahren schwieriger zu entdecken sind als negative. Da nicht alle Persönlichkeitseigenschaften bzw. -skalen in gleichem Maße durch Antwortverzerrungen beeinflusst werden, untersuchten Haywood, Grossman und Hardy (1993) den Zusammenhang zwischen Antwortverhalten und den verschiedenen Skalen eines Persönlichkeitsfragebogens zur Erfassung der Big Five anhand einer Stichprobe von Männern, die wegen eines Sexualdelikts beschuldigt und zur psychiatrisch-psychologischen Begutachtung überwiesen wurden. Von den 16 Skalen des Fragebogens waren sieben nicht mit positiven Antwortverzerrungen verknüpft. Signifikante positive Korrelation fanden sich mit Skalen, die Probanden als emotional stabil, sorglos, selbstsicher und kontrolliert charakterisieren, signifikante negative Korrelationen mit den Skalen, die Probanden als misstrauisch, gerissen, besorgt, selbstgenügsam und ängstlich charakterisieren.

Dass positive Antwortverzerrungen nicht nur auf die erreichten Skalenwerte der Verfahren (dimensionale Auswertung), sondern auch auf die kategoriale Auswertung Einfluss haben, konnte ebenfalls gezeigt werden. In der Studie von Fiduccia (2011) wurde einer Stichprobe stationär-psychiatrischer Patienten mit Doppeldiagnose zunächst mit Standard-Instruktion, später mit Fake-good-Instruktion ein Selbstbeurteilungsinstrument zu Persönlichkeitsstörungen vorgelegt und auch ein semi-strukturiertes Interview mit ihnen durchgeführt. Nach Fake-Instruktion erzielten die Probanden vor allem im Fragebogen, aber auch im Interview um rund die Hälfte reduzierte Werte. Sie waren also in der Lage, klinische Persönlichkeitsmerkmale zu minimieren. Über 70% der Probanden, bei denen zunächst eine Persönlichkeitsstörung diagnostiziert worden war, hatten nach Fake-Instruktion keine Achse-II-Diagnose mehr. Der Effekt war bei einigen Persönlichkeitsstörungen, die durch mehr sozial unerwünschte Merkmale charakterisiert sind (Antisoziale und Borderline Persönlichkeitsstörung), stärker ausgeprägt als bei anderen (Schizoide, Histrionische, Ängstlich-vermeidende, Abhängige und Zwanghafte Persönlichkeitsstörung). Myerholtz und Rosenberg (1997) untersuchten mit ähnlichem Design (Standard- vs. Fake-Instruktion) Personen, die unter Alkoholeinfluss ein Fahrzeug geführt hatten und im Rahmen einer gerichtlichen

Auflage an einem Wochenendprogramm teilnahmen. Von den Probanden, die in einem Screeningverfahren zu körperlicher Substanzabhängigkeit zunächst auffällige Werte erzielt hatten, erreichten 71% in der Fake-Bedingung unauffällige Werte.

Positive Antwortverzerrungen führen also zu veränderten Werten auf Skalen, die Psychopathologie und abweichende bzw. sozial (un-)erwünschte Persönlichkeitsmerkmale erfassen. Sie verzerren auch die kategoriale Testauswertung bzw. die Klassifikation und können somit den Befund von Verfahren, die auf Selbstauskünften beruhen, erheblich verfälschen. Deshalb wird immer wieder betont, dass bei forensischen Begutachtungen das Erfassen und Berücksichtigen des Antwortverhaltens bei der Interpretation der Befunde unabdingbar ist (Otto, 2008). Empirische Studien an forensischen Populationen, die zeigen, dass das Kontrollieren von positiven Antwortverzerrungen bzw. eine Korrektur der Testergebnisse zu einer Verbesserung der Validität führt, gibt es jedoch kaum. Vielmehr weisen einige Studien darauf hin, dass das Auspartialisieren von sozialer Erwünschtheit die Zusammenhänge zwischen Selbstbericht und ausgewählten Kriteriumsvariablen (z.B. antisoziales Verhalten, Rückfälligkeit) sogar reduzieren kann (Übersicht bei Tan & Grace, 2008). Die Ergebnisse der o.g. Studien sagen auch nichts darüber aus, wie im individuellen Fall Befunde der eingesetzten Verfahren zu interpretieren sind bzw. welche klinische Relevanz den Antwortverzerrungen zukommt (Rogers, 2008c). Zwar wurden allgemeine Empfehlungen formuliert, wie Antwortverhalten im forensischen Kontext erfasst werden kann (z.B. Otto, 2008), Einigkeit bzw. empirisch fundierte Empfehlungen, mit welchen spezifischen Methoden welche Antwortverzerrungen am besten zu erfassen sind, gibt es jedoch nicht.

2.3 Methoden zur Erfassung von Antwortverzerrungen

Nach Rogers (2008b) muss eine Methode zur Aufdeckung von Antwortverzerrungen standardisiert sein, eine konzeptionelle Basis besitzen, empirisch validiert sein und zuverlässig das Antwortverhalten klassifizieren bzw. eine spezifische Antwortverzerrung erfassen können. Er gibt in seinem Standardwerk einen Überblick über verschiedene Strategien zur Erfassung negativer und positiver Antwortverzerrungen, die er in zwei Kategorien einteilt: unlikely presentations und amplified presentations. Erstere fokussieren auf das Berichten von ungewöhnlichen oder atypischen Merkmalen, letztere auf die Häufigkeit oder Intensität der berichteten Merkmale, jeweils im Vergleich zur Referenzpopulation, z.B. Patienten mit entsprechender psychischer Störung. Auf die Erfassung von negativen Antwortverzer-

rungen soll hier nicht näher eingegangen werden, eine Übersicht über deutschsprachige Instrumente zur Aufdeckung von Aggravation oder Simulation sowie suboptimalem Leistungsverhalten findet sich bei Merten (2014). Für positive Antwortverzerrungen identifiziert Rogers Strategien, die alle den amplified presentations zuzuordnen sind. Sie zielen darauf ab, dass Personen mit positiven Antwortverzerrungen auch kleinere Schwächen oder Fehler oder verbreitete Beschwerden nicht eingestehen, positive Eigenschaften übertreiben oder Symptome, die für bestimmte klinische Populationen typisch sind, verleugnen. Es wurde eine Vielzahl von Skalen entwickelt, die verschiedene Formen von positiven Antwortverzerrungen erfassen (Übersicht z.B. bei Rogers, 2008b; c). Manche wurden speziell für mehrdimensionale Persönlichkeitsinventare entwickelt und in diese integriert (Übersicht für das MMPI bei Greene, 2008; Übersicht für gängige Persönlichkeitsinventare bei Sellbom & Bagby, 2008). Allerdings wird deren Einsatz aufgrund z.T. unbefriedigender Gütekriterien kontrovers diskutiert (Rogers, 2008b; Littmann, 2007). Baer und Miller (2002) untersuchten in einer Metaanalyse die Validität der MMPI-Skalen zu positiven Antwortverzerrungen. Dabei zeigte sich, dass durch die 12 verschiedenen Skalen zwischen 69% und 84% der Probanden mit Fake-good-Verhalten und zwischen 67% und 88% der offen antwortenden Probanden korrekt identifiziert wurden. Zwischen 53% und 75% der Probanden mit auffälligen Skalenwerten hatten ihre Antworten verzerrt, zwischen 87% und 92% der Probanden mit unauffälligen Skalenwerten hatten offen geantwortet. Die diagnostische Validität (s. Kap. 2.4) war noch geringer, wenn Probanden untersucht wurden, die zuvor darin gecoacht worden waren, in Kontrollskalen unauffällig abzuschneiden. Die Befunde zur Verbesserung der Validität durch den Einsatz mehrerer Validitätsskalen waren uneinheitlich. In der Untersuchung von Edens et al. (2001) differenzierten zwar die Kontrollskala des PPI und eine Skala zu sozialer Erwünschtheit signifikant zwischen instruierten und ehrlich antwortenden Kontrollprobanden, die diagnostische Validität war jedoch vor allem für die PPI-Kontrollskala gering und es kam zu einem beträchtlichen Anteil an Fehlklassifikationen.

In vielen Instrumenten, die bei forensischen Begutachtungen zum Einsatz kommen, sind jedoch gar keine Kontrollskalen integriert. Um bei testpsychologischen Untersuchungen dennoch Verfälschungstendenzen aufdecken zu können, kann man auf Verfahren zurückgreifen, die ausschließlich entwickelt wurden, um Antwortverzerrungen zu erfassen. Solche existieren jedoch vorwiegend für Aggravation und Simulation. Bei deren Anwendung ist zudem zu berücksichtigen, dass es nicht unproblematisch ist anzunehmen, dass sich der Antwortstil bei einem Verfahren direkt auf andere Verfahren übertragen lässt. Bei Otto

(2008) findet man eine Reihe von Beispielen, wie Personen in Bezug auf unterschiedliche Merkmale unterschiedliche Formen von Antwortverzerrungen zeigen können. Er geht davon aus, dass hybrides Antwortverhalten im forensischen Kontext sogar am häufigsten zu finden ist.

Aus diesen Gründen sollte auch die Erfassung von Antwortverzerrungen multimodal, durch den Einsatz verschiedener Skalen zu unterschiedlichen Antwortverzerrungen, erfolgen.

2.4 Validierung von Verfahren zur Erfassung von Antwortverzerrungen

2.4.1 Validitätsindizes

Die Validität eines Tests beschreibt, wie gut ein Test das misst, was er messen soll. Sie wird unterteilt in die Inhaltsvalidität, die vor allem bei der Konstruktion eines Instruments kritisch reflektiert werden muss, aber kaum empirisch überprüft bzw. quantifiziert werden kann (Bühner, 2011), die Kriteriums- und die Konstruktvalidität.

Das wichtigste Gütekriterium eines Fragebogens zur Erfassung von Antwortverzerrungen ist die Kriteriumsvalidität, genauer die diagnostische Validität. Das Antwortverhalten soll durch das Instrument möglichst treffend in unauffällig oder auffällig eingestuft werden. Um die diagnostische Validität zu bestimmen, wird die Klassifikationsübereinstimmung zwischen dem neuen Testverfahren und einem Referenzstandard, dem so genannten „Goldstandard“ (s. Kap. 2.4.2) überprüft. Das Verfahren kann als valide angesehen werden, wenn die Rate der Fehlklassifikationen gering ist. Es müssen zwei Arten von Fehlklassifikationen unterschieden werden: Personen, die ehrlich antworten, aber ein auffälliges Testergebnis erzielen (falsch Positive), und Personen, die nicht aufrichtig antworten, aber ein unauffälliges Testergebnis erzielen (falsch Negative). Das Ausmaß an korrekten bzw. Fehlklassifikationen wird durch folgende Güteindizes quantifiziert:

- Sensitivität: prozentualer Anteil der unaufrichtig antwortenden Probanden mit auffälligem Testbefund
- Spezifität: prozentualer Anteil der aufrichtig antwortenden Probanden mit unauffälligem Testbefund
- positiver Vorhersagewert: prozentualer Anteil der korrekt erkannten unaufrichtig antwortenden Probanden an der Anzahl aller Probanden mit auffälligem Testbefund
- negativer Vorhersagewert: prozentualer Anteil der korrekt erkannten aufrichtig antwortenden Probanden an der Anzahl aller Probanden mit unauffälligem Testbefund

- Gesamttrefferquote: prozentualer Anteil der korrekt als aufrichtig oder unaufrichtig antwortend eingestuften Probanden an allen untersuchten Probanden.

Die Grundlage dieser Überprüfung bildet die kategoriale Auswertung des Instruments. Positive Antwortverzerrungen sind jedoch häufiger als dimensionale Konstrukte konzipiert (Rogers, 2008d), d.h. die Neigung, sich selbst als möglichst positiv oder unauffällig zu präsentieren, ist mehr oder weniger stark ausgeprägt anstatt vorhanden oder nicht vorhanden. Um dieser Konzeptualisierung gerecht zu werden, ist auch die dimensionale, also korrelative Übereinstimmung des neu entwickelten Instruments mit bereits etablierten Instrumenten, die ähnliche Antwortverzerrungen erfassen, zu bestimmen (konvergente Validität). Diese ist Teil der Konstruktvalidierung. Während die Korrelationen mit Verfahren des gleichen Gültigkeitsbereichs hoch sein sollten, sollten die Korrelationen mit Tests anderer Gültigkeitsbereiche gering sein (diskriminante Validität), um sicherzustellen, dass das Verfahren nur das interessierende Konstrukt und keine anderen Merkmale erfasst. Allerdings ist diese Forderung bei Skalen zu Antwortverzerrungen oft nicht erfüllt, es zeigen sich immer wieder Zusammenhänge z.B. mit verschiedenen überdauernden Persönlichkeitsmerkmalen (s. Kap. 2.2.2.2). Zudem muss im Rahmen einer Validierung überprüft werden, ob die Ergebnisse von Skalen, die eine Dissimulation von Symptomen oder Beschwerden erfassen, nicht auch durch negative Antwortverzerrungen beeinflusst werden. Denn es ist möglich, dass Personen mit positiven Antwortverzerrungen durch das Verneinen oder Minimieren von Beschwerden auffallend niedrige Skalenwerte erzielen, Personen, die sich als besonders auffällig oder krank präsentieren möchten, auf der Skala dagegen mehr oder intensivere Beschwerden angeben, als tatsächlich vorhanden sind, und damit auffallend hohe Werte erzielen. Dann ist ggf. die zweistufige Interpretation des Skalenwerts (unauffällig vs. dissimulierend) durch eine dreistufige (unauffällig vs. dissimulierend vs. aggravierend) zu ersetzen. Dies ist zum Beispiel bei den Kontrollskalen des MMPI der Fall.

Zur Konstruktvalidierung gehört auch die Überprüfung der Zusammenhangsstruktur der einzelnen Items (Hartig, Frey & Jude, 2012). Um eine angenommene Dimensionsstruktur und damit die faktorielle Validität eines Instruments zu prüfen, muss eine konfirmatorische Faktorenanalyse berechnet werden. Wenn es jedoch nur darum geht, Zusammenhänge zwischen Items deskriptiv zu beschreiben bzw. miteinander korrelierende Items zu Subskalen zusammenzufassen, um z.B. die (Dis-)Simulation bestimmter Symptombereiche zu erfassen, ist eine exploratorische Faktorenanalyse ausreichend. Diese findet man in Validierungsstudien auch weitaus häufiger.

Ferner sollten die Auswirkungen der gezeigten Antwortverzerrungen auf weitere (testpsychologische) Befunde der Begutachtung untersucht werden. Die Überprüfung der praktischen Relevanz auffälliger Befunde bei Instrumenten von Antwortverzerrungen kann nach Hartig et al. (2012) nicht nur unter dem Gesichtspunkt der Kriteriums-, sondern auch unter dem der Konstruktvalidierung gesehen werden. Denn wenn sich theoretisch fundierte Hypothesen über den Zusammenhang zwischen den Testbefunden und bestimmten Außenkriterien bestätigen lassen, kann das zum einen als Beleg dafür dienen, dass die Interpretation des Tests valide ist und wirklich spezifische Antwortverzerrungen erfasst werden (Konstruktvalidität). Zum anderen kann damit gezeigt werden, dass das Verfahren seinen Zweck erfüllt, nämlich dass die Angaben und Selbsteinschätzungen von Probanden mit auffälligen Testwerten tatsächlich nur unter Vorbehalt interpretiert werden können oder gar auf den Einsatz von Selbstbeurteilungsverfahren bei der Begutachtung verzichtet wird.

Da die Reliabilität nicht nur zu den Hauptgütekriterien eines Testverfahrens gehört, sondern eine mangelnde Reliabilität, d.h. ein hoher Anteil an unsystematischer oder Fehlervarianz bei der Erfassung eines Merkmals, Grund für eine niedrige Kriteriumsvalidität sein kann (Bühner, 2011), sollte diese vor Berechnung der Validitätsindizes geprüft werden.

2.4.2 Untersuchungsdesign

Um die diagnostische Validität von Verfahren zur Erfassung von Antwortverzerrungen zu bestimmen, müssen die Ergebnisse von zwei Gruppen miteinander verglichen werden: Personen, die offen antworten, und Personen, die Antwortverzerrungen zeigen. Um die Probanden einer der Gruppen zuordnen zu können, muss jedoch das tatsächliche Antwortverhalten bekannt bzw. ein zuverlässiges Kriterium für die Gruppenzuordnung vorhanden sein. Im Idealfall werden die Personen unter realen Bedingungen untersucht, d.h. dass z.B. für die Validierung Daten von Begutachtungsprobanden zur Verfügung stehen. Dieses Vorgehen des sog. Known-groups comparison (Rogers, 2008d) setzt voraus, dass das tatsächliche Antwortverhalten von Probanden bekannt ist. Dazu benötigt man ein valides Referenzkriterium, der sog. Goldstandard, der eine Gruppeneinteilung unabhängig von dem zu validierenden Instrument ermöglicht (Schmand & Merten, 2009). Ein solcher Goldstandard kann entweder ein weiteres testpsychologisches Verfahren sein oder ein Expertenurteil bzw. eine klinische Einschätzung.

Die Entwicklung eines spezifischen testpsychologischen Verfahrens wäre aber überflüssig, wenn bereits ein valides Testverfahren existiert, das als Goldstandard herangezogen werden kann. Die SN-S wurde konstruiert, da kein Verfahren verfügbar war, das Dissimulati-

onstendenzen i.S. einer Beschwerdenminimierung bei Begutachtungsprobanden erfasst. Als Referenzkriterium können in einem solchen Fall lediglich Skalen mit einem ähnlichen Messanspruch herangezogen werden. Da diese jedoch auch keine zu 100 Prozent zuverlässige Einstufung ermöglichen, fallen die Schätzungen für die Güteindizes des zu validierenden Instruments unterschiedlich aus, je nachdem, welches Instrument als Goldstandard herangezogen wird.

Als Referenzstandard bzw. Expertenurteil bzgl. des Vorliegens von Antwortverzerrungen kann der Vergleich zwischen den Selbstangaben des Probanden mit verschiedenen Außenkriterien, wie einem Expertenurteil oder den Befunden objektiver Verfahren, dienen. So kann z.B. eine Minimierung von Substanzkonsum festgestellt werden, indem die Übereinstimmung zwischen berichteter Konsummenge einer entsprechenden Substanz und den Ergebnissen einer Haaranalyse geprüft wird. Die Minimierung devianter sexueller Interessen könnte durch den Vergleich der Selbstangaben mit den Ergebnissen indirekter Verfahren (z.B. Impliziter Assoziationstest zur Erfassung pädosexueller Neigungen, Yundina & Nedopil, 2010) festgestellt werden. Für die meisten bei forensischen Begutachtungen erfassten Merkmale sind jedoch entsprechende Methoden nicht verfügbar. Eine andere Möglichkeit, um für Validierungsstudien einen Referenzstandard zu konstruieren, ist, i.S. einer Diskrepanzdiagnostik, der Vergleich der Selbstangaben von Probanden bezüglich fragestellungsrelevanter Variablen (z.B. dem Vorliegen einer Persönlichkeitsstörung) mit der klinischen Einschätzung durch einen geschulten Beurteiler. Allerdings hängt die Validität des Referenzstandards von der Güte des Expertenurteils ab. Zudem ist dieses Vorgehen in der Praxis für Validierungsstudien kaum zu leisten, vor allem für Fragebögen wie die SN-S, mit der unterschiedlichste Auffälligkeiten und Beschwerden erfasst werden, und dementsprechend viele Merkmale durch den Experten beurteilt werden müssten.

In Validierungsstudien für Instrumente von negativen Antwortverzerrungen wurde daher bisweilen auf eine globale klinische Einschätzung des Antwortverhaltens als Referenzstandard zurückgegriffen. Allerdings gibt es für positive Antwortverzerrungen bislang weit weniger empirisch gesicherte Kriterien und Methoden, die einer solchen klinischen Einschätzung zugrunde gelegt werden können (Rogers, 2008b). Bei Begutachtungen könnte zwar die Beurteilung eines erfahrenen psychiatrischen Sachverständigen als Referenzstandard dienen, allerdings nur dann, wenn er diesbezüglich eine Einschätzung vornimmt, ohne die Ergebnisse der testpsychologischen Zusatzbegutachtung zu kennen. Dies ist jedoch in der Praxis nicht umsetzbar. Die testpsychologischen Zusatzbegutachtungen werden gerade deshalb durchgeführt, weil eine zuverlässige klinische Einschätzung des Antwortverhaltens

schwierig ist, bzw. um die testpsychologischen Ergebnisse bei der psychiatrischen Gesamtbeurteilung nutzen zu können.

Da es also für die Erfassung von Antwortverzerrungen bislang keinen echten Goldstandard gibt, der einen Vergleich bekannter Gruppen ermöglicht, wird in Validierungsstudien häufig auf experimentelle Designs zurückgegriffen (Rogers, 2008d). Dabei werden Kontrollprobanden gezielt instruiert, zu simulieren oder dissimulieren oder aber ehrlich zu antworten bzw. sich anzustrengen und ihre bestmöglichen Leistungen zu zeigen. Die Antworten bzw. Leistungen können dann mit denen von ehrlich antwortenden Kontrollprobanden oder aber, wenn es um negative Antwortverzerrungen geht, idealerweise mit denen von realen Patienten verglichen werden. Die externe Validität dieses Designs ist gegenüber dem Vergleich bekannter Gruppen eingeschränkt. Denn Begutachtungsprobanden unterscheiden sich bezüglich einer Vielzahl von Merkmalen von gesunden Kontrollprobanden, deren Einfluss auf das Antwortverhalten aber bislang kaum oder gar nicht untersucht wurde. So handelt es sich bei Begutachtungsprobanden häufig um Personen mit psychischen Symptomen oder Störungen, während die Mehrheit der Kontrollprobanden psychisch gesund ist. Begutachtungsprobanden bringen außerdem in die Untersuchung auch eine ganz spezifische Motivation mit, das Ergebnis der Untersuchung besitzt für sie – im Gegensatz zu den Kontrollprobanden – eine hohe persönliche Relevanz. Daher ist es wichtig, die Instruktion so zu formulieren, dass bei den Kontrollprobanden eine hohe Motivation bzw. ein hohes Engagement bei der Testbearbeitung entstehen, z.B. indem die persönliche Relevanz des Forschungsprojekts vermittelt wird (Rogers, 2008d).

Die Wahl des Untersuchungsdesigns und der eingesetzten Methoden beeinflusst, wie genau und unverzerrt die Validität eines Verfahrens bestimmt werden kann. Je zuverlässiger der Referenzstandard eine Gruppeneinteilung ermöglicht bzw. je besser das Antwortverhalten experimentell manipuliert wird, desto weniger ist die Validitätsschätzung verzerrt.

Ein gravierendes Problem in experimentellen Studien zu positiven Antwortverzerrungen ist, dass die Personen der Kontrollgruppe auch unaufrichtig oder sozial erwünscht antworten bzw. dissimulieren können. Dies führt bei der Bestimmung der Validitätsindizes zu einer erhöhten Zahl falsch positiver Treffer und damit zu einer Abnahme der Spezifität und des positiven Vorhersagewerts. Gelingt dagegen die Manipulation durch die Fake-Instruktion nicht, antworten also die Probanden der Fake-Bedingung nicht sozial erwünscht oder dissimulierend, steigt die Anzahl falsch negativer Treffer, die Sensitivität und der negative Vorhersagewert sinken.

Studien mit experimentellen Designs liefern daher lediglich eine Schätzung der Güte des Verfahrens. Die Ähnlichkeit mit der praktischen Anwendungssituation sollte so groß wie möglich sein, um zu möglichst unverzerrten Schätzungen zu kommen (Schmand & Mer-ten, 2009). Die möglichen Einflussfaktoren, die zu einer Verzerrung der diagnostischen Validität führen können, müssen bei der Wahl der Untersuchungsmethoden, insbesondere bei der Stichprobenrekrutierung und der Formulierung der Instruktionen, kritisch reflektiert und bei der Interpretation der Untersuchungsergebnisse berücksichtigt werden.

2.5 Die Supernormalität-Skala (SN-S)

Die Supernormalität-Skala wurde von Cima und Kollegen (2003a, 2003c) entwickelt, um das in Kapitel 2.1 erläuterte Konstrukt „Supernormalität“ zu erfassen. Die ursprüngliche Fassung (Supernormality Scale – SS) bestand aus 44 (Cima et al., 2003a) bzw. nach einer ersten Revision aus 37 Items (Cima et al., 2003c) mit einer 2-stufigen Antwortskala im Ja-Nein-Format. Höhere Werte sprechen für eine stärkere Neigung zu Supernormalität. Elf Items waren aus dem MMPI-2 und der Social Desirability Scale (Crowne & Marlowe, 1960) entnommen, die übrigen wurden von erfahrenen forensischen Psychiatern und Psychologen formuliert. Außerdem wurden fünf Dummy-Items mit neutralem Inhalt eingefügt, um den Testzweck zu verschleiern. Diese ersten Fassungen wurden an Stichproben aus ehrlich antwortenden, gesunden Kontrollprobanden, inhaftierten Straftätern und instruierten Kontrollprobanden, Patienten der Allgemeinpsychiatrie und Maßregelvollzugspatienten überprüft. Eine Hauptkomponentenanalyse der 37-Item-Version zeigte eine zweifaktorielle Struktur mit den Faktoren Supernormalität und soziale Erwünschtheit. Die interne Konsistenz war mit Cronbachs $\alpha = .86$ zufriedenstellend, die Retest-Reliabilität nach sechs Wochen mit $r_{tt} = .90$ hoch. Die einzelnen Stichproben erzielten erwartungsgemäß signifikant unterschiedliche Mittelwerte, auch nach Kontrolle von Alter und Geschlecht der Probanden: Psychiatrische Patienten ohne kriminelle Vorgeschichte ($n = 28$, $M = 10.18$) neigten weniger zu Supernormalität als alle anderen Untersuchungsgruppen. Die Maßregelvollzugspatienten ($n = 59$, $M = 18.15$) hatten signifikant höhere Werte, zeigten also mehr supernormales Antwortverhalten als ehrlich antwortende Kontrollprobanden ($n = 45$, $M = 15.33$), sie unterschieden sich aber nicht signifikant von psychisch gesunden Straftätern ($n = 49$, $M = 15.67$) und den instruierten Kontrollprobanden ($n = 50$, $M = 18.48$). Die ehrlich antwortenden Kontrollprobanden erzielten außerdem signifikant niedrigere Werte als die instruierten Kontrollprobanden.

Die 44-Item-SS korrelierte signifikant positiv mit der Sozialen Erwünschtheitsskala des MMPI-2 ($r = .48$) und signifikant, aber gering negativ mit der Simulationsskala des MMPI-2 (F-Skala; $r = -.20$). Dass sie nicht mit der Dissimulationsskala (K-Skala) des MMPI-2 korrelierte, werteten die Autoren als Zeichen dafür, dass es sich bei Supernormalität um ein von Dissimulation unabhängiges Konstrukt handelt. Diese Abgrenzung wurde jedoch, wie bereits erwähnt, für die späteren Fassungen des Instruments aufgegeben, ohne dass dies von den Autoren näher erläutert bzw. begründet wurde.

Mit der 37-Item-Fassung konnten außerdem für forensische Patienten moderate, aber signifikante Zusammenhänge zwischen Supernormalität und der Minimierung von Verantwortung für das begangene Delikt gezeigt werden. Eine kleine Substichprobe der Maßregelvollzugspatienten ($n = 12$), die alle wegen sexuellen Missbrauchs verurteilt worden waren, nahm außerdem an einem Experiment zur Gedankenunterdrückung teil. Sie wurden im ersten Durchgang instruiert, nicht an einen weißen Bären zu denken, im zweiten Durchgang, an alles denken zu dürfen, auch an weiße Bären. Im dritten Durchgang wurden sie instruiert, nicht an Kinder zu denken, im vierten Durchgang lautete die Anweisung, an alles denken zu dürfen, auch an Kinder. Das Auftreten des Zielgedankens sollte jeweils notiert werden. Für die Bedingung Weißer Bär zeigte sich der typische Reboundeffekt (signifikant häufigeres Auftreten des Zielgedankens in Durchgang 2), für die Bedingung Kinder jedoch nicht. Während die SS-Ergebnisse nicht mit der Auftretenshäufigkeit des Zielgedankens Weißer Bär in Durchgang 2 korrelierten, zeigte sich ein signifikanter negativer Zusammenhang ($r = -.59$, $p < .05$) mit der berichteten Häufigkeit des Zielgedankens Kinder in Durchgang 4, d.h. je stärker die Neigung zu supernormalem Antwortverhalten ausgeprägt war, desto geringer war die Bereitschaft, Gedanken an Kinder zu berichten.

Für die 37-Item-SS wurde die diagnostische Validität anhand des Vergleichs ehrlich antwortender und instruierter Kontrollprobanden bestimmt. Die Sensitivität (74%) und vor allem die Spezifität (42%) waren nicht zufriedenstellend. Dies führte zu einer Überarbeitung der Skala (Cima et al., 2008). Einige Items wurden entfernt und stattdessen Items aus der Dissociative Experience Scale (Bernstein & Putnam, 1986) und der Revised Hallucination Scale (Morrison, Wells & Nothard, 2002) hinzugefügt. Diese revidierte Fassung (Supernormality Scale-Revised – SS-R) besteht aus 34 Items, die sieben verschiedenen Bereichen zugeordnet werden können: soziale Erwünschtheit (4 Items), affektive Symptome (7 Items), Zwangssymptome (4 Items), psychotische Symptome (6 Items), dissoziative Symptome (5 Items), Aggression (7 Items) und Angst (1 Item), sowie 16 Dummy-Items. Die Items sind auf einer 4-stufigen Skala einzuschätzen. Der Skalenwert entspricht dem Sum-

menwert über alle 34 Items, wobei zwei Items vor der Auswertung umgepolt werden müssen, und kann zwischen 34 und 136 Punkten liegen. Kleinere Werte sprechen für eine höhere Tendenz zu Supernormalität. Die Skala wurde an einer Stichprobe ($N = 135$) aus forensischen Patienten, allgemeinspsychiatrischen Patienten, ehrlich antwortenden Studenten und instruierten Studenten überprüft (Cima et al., 2008). Die Retest-Reliabilität war ausreichend ($r_{tt} = .79$), die interne Konsistenz gut (Cronbachs $\alpha = .88$). Die SS-R-Werte korrelierten sowohl in der Gesamtstichprobe als auch in den Substichproben signifikant mit den Ergebnissen einer Paranoia-Skala ($r = .44-.71$), deren Items dem MMPI entnommen waren, d.h. je größer die Tendenz zu supernormalem Antwortverhalten war, desto weniger Symptome übermäßigen Misstrauens wurden berichtet. Forensische Patienten ($n = 63$, $M = 61.32$) und instruierte Studenten ($n = 20$, $M = 56.65$) hatten signifikant niedrigere SS-R-Mittelwerte als ehrlich antwortende Studenten ($n = 26$, $M = 67.58$) und Allgemeinspsychiatriepatienten ($n = 26$, $M = 66.27$). Bei einem Cut-off-Wert von ≤ 60 Punkten zeigte die Skala eine gute diagnostische Validität mit einer Sensitivität von 80%, einer Spezifität von 92%, einem positiven Vorhersagewert von 86% und einem negativen Vorhersagewert von 86%. Cut-off-Werte und Untersuchungen zur diagnostischen Validität der Subskalen wurden bislang nicht publiziert.

In einer weiteren Validierungsstudie (Cima et al., 2008) an gesunden Kontrollprobanden ($n = 118$) und Maßregelvollzugspatienten ($n = 34$) wies die SS-R ebenfalls eine hohe Retest-Reliabilität ($r_{tt} = .86$) und eine akzeptable interne Konsistenz (Cronbachs $\alpha = .76$) auf. Die SS-R-Werte korrelierten negativ mit den Werten des PPI ($r = .56$).

Cima et al. (2009) untersuchten diesen Zusammenhang genauer an einer Stichprobe Tatverdächtiger, die sich in Untersuchungshaft befanden ($n = 39$), und inhaftierter verurteilter Straftäter ($n = 36$). Es wurde angenommen, dass supernormales Antwortverhalten sowohl vom Untersuchungskontext als auch von der Persönlichkeit abhängig ist. Diese Hypothese konnte jedoch nicht bestätigt werden, sondern es zeigte sich erneut ein negativer Zusammenhang zwischen Psychopathie, gemessen mit dem PPI, und Supernormalität und zwar unabhängig vom Untersuchungskontext. Dagegen konnte für den Zusammenhang zwischen Simulation, Psychopathie und Untersuchungskontext die Hypothese bestätigt werden. Probanden mit hohen Psychopathiewerten im PPI zeigten eine stärkere Simulationsneigung als Probanden mit niedrigen Psychopathiewerten, aber nur in der Untersuchung vor Urteilsspruch. Bei bereits verurteilten Straftätern konnte kein Zusammenhang zwischen Psychopathie und Simulation gefunden werden. Eine Diskussion dieser Befunde

erfolgt unten im Zusammenhang mit Ergebnissen der revidierten deutschen Fassung der SS-R.

Diese wurde 2009 in der Abteilung für Forensische Psychiatrie der LMU übersetzt. Die Gütekriterien dieser ersten Fassung wurden an einer Stichprobe von 72 Probanden (25 Begutachtungsprobanden, 17 Patienten in stationär-psychiatrischer Behandlung und 30 Studenten) untersucht (Veress, 2010). Die interne Konsistenz war mit Cronbachs $\alpha = .85$ als gut zu beurteilen. Die mittlere Inter-Item-Korrelation war jedoch niedrig ($MIC = .16$) und die part-whole-korrigierten Trennschärfen fielen für viele Items zu gering aus. Dennoch konnte die Hypothese bestätigt werden, dass sozialmedizinische Begutachtungsprobanden ($n = 11$, $M = 66.6$) ein geringeres Ausmaß an Supernormalität zeigen als strafrechtliche Begutachtungsprobanden ($n = 11$, $M = 56.3$). Sie unterschieden sich auch signifikant von der studentischen Kontrollstichprobe ($n = 27$, $M = 58.4$). Die Hypothese, dass Probanden, bei denen eine Rückfallprognose erfragt war ($n = 8$, $M = 55.6$), mehr zu supernormalem Antworten neigen als Probanden, die zur Frage der Schuldfähigkeit begutachtet wurden ($n = 5$, $M = 57.5$), konnte anhand der kleinen Stichprobe dagegen nicht bestätigt werden. Die SS-R-Werte korrelierten erwartungsgemäß signifikant ($\tau = .24$, $p < .01$) mit der Offenheitsskala des Freiburger Persönlichkeitsinventars (Fahrenberg, Hampel & Selg, 2010). Es zeigte sich kein Zusammenhang zwischen Supernormalität und Intelligenz.

Aufgrund der ersten Befunde zur deutschsprachigen SS-R wurde die Übersetzung einiger Items überarbeitet. Insbesondere die beiden ursprünglich negativ gepolten Items („Ich neige nicht zu Zwangshandlungen“ und „Im Augenblick gibt es keine Situation, in der ich aggressiv reagieren würde“) wurden in positiv gepolte Aussagen umformuliert. Darüber hinaus wurden die Formulierungen der Antwortskala modifiziert (ursprünglich „nie“, „manchmal“, „oft“, „immer“), da sie von vielen Probanden als mit einigen Aussagen inkompatibel eingeschätzt wurden. In der aktuellen Fassung lauten die entsprechenden Antwortkategorien „trifft gar nicht zu“, „trifft eher nicht zu“, „trifft eher zu“, „trifft absolut zu“. Für die ursprüngliche Übersetzung war nach der Formel von Jacobson & Truax (1991) ein Cut-off-Wert von 61 Punkten errechnet worden, der für die revidierte Fassung beibehalten wurde, es wurde jedoch ein neuer Name (SNS) gewählt. Nach Rücksprache mit den Originalautoren wurde eine Modifikation weiterer Itemformulierungen vorgenommen. Bei der aktuellen Fassung der SN-S handelt es sich um die autorisierte Übersetzung der SS-R (Veress, Merten & Yundina, 2010; s. Anhang A2).

Mit dieser Fassung wurde zunächst noch einmal an einer kleinen Stichprobe von Begutachtungsprobanden die Hypothese überprüft, dass Schuldfähigkeits- und Prognoseprobanden

sich hinsichtlich der Auftretenshäufigkeit positiver Antwortverzerrungen unterscheiden (Tippelt et al., 2010). Im Gegensatz zu den früheren Studien sollten nun die kategorialen Unterschiede überprüft werden. Denn bei der SN-S handelt es sich um ein Screening-Instrument, das bei Begutachtungen dazu dienen soll, die Probanden zu identifizieren, bei denen aufgrund positiver Antwortverzerrungen die Befunde von Selbstauskunftsverfahren nur unter Vorbehalt bzw. nicht mehr zuverlässig zu interpretieren sind. Demnach sollte der Schwerpunkt bei den weiteren Validierungsstudien auf die dichotome Auswertung der SN-S gelegt werden. Analysiert wurden die SN-S-Werte von Probanden, die in der Abteilung für Forensische Psychiatrie der LMU im Jahr 2010 zur Frage der Rückfallprognose ($n = 10$) oder zur Frage der Schuldfähigkeit ($n = 15$) begutachtet worden waren. Entsprechend der Hypothese zeigten alle der untersuchten Prognoseprobanden, aber nur 6 der 15 Schuldfähigkeitsprobanden eine Dissimulationsneigung ($p < .01$).

Bantel (2011) untersuchte anhand einer Stichprobe von ehrlich antwortenden ($n = 28$) und instruierten Kontrollprobanden ($n = 28$) sowie Begutachtungsprobanden der Abteilung für Forensische Psychiatrie der LMU ($n = 28$) die Zusammenhänge zwischen Supernormalität, Intelligenz und Psychopathie. Die Kontrollprobanden waren erwachsene Personen ohne psychiatrische und strafrechtliche Vorgeschichte, die aus dem persönlichen Umfeld des Autors rekrutiert wurden, wobei darauf geachtet wurde, ein hinsichtlich Alter, Geschlecht und Bildungshintergrund heterogenes Kollektiv zusammenzustellen. Den Probanden wurde die SN-S, das PPI-R (Alpers & Eisenbarth, 2008) sowie das Intelligenzscreeningverfahren Wortschatztest (WST; Schmidt & Metzler, 1992) mit einer Instruktion zur Bearbeitung der Instrumente und einem adressierten und frankierten Rückumschlag per Post zugesendet. Die Rücksendung der Instrumente erfolgte anonym. Von den 110 versandten Umschlägen wurden 75% zurückgesendet, nach Ausschluss von unvollständigen Unterlagen konnten 56 Kontrollprobanden in die Analyse eingeschlossen werden. Die Daten der Begutachtungsprobanden wurden den Akten entnommen. Es handelte sich um Personen (89% männlich), die zur Frage der Rückfallprognose oder zur Frage der Schuldfähigkeit begutachtet worden waren. Eine getrennte Auswertung für unterschiedliche Begutachtungsfragestellungen konnte aufgrund des geringen Stichprobenumfangs nicht vorgenommen werden. Der WST lag von allen Probanden vor, das PPI-R von 14 Probanden. Die Hypothese, dass sich die Stichproben signifikant hinsichtlich des Ausmaßes an positiven Antwortverzerrungen unterscheiden, konnte nicht bestätigt werden. Beim Vergleich der SN-S-Werte zeigten sich im Gegensatz zu den Ergebnissen von Tippelt et al. (2010) weder signifikante Mittelwertunterschiede (Begutachtungsprobanden: $M = 66.8$; Kontrollprobanden: $M = 68.0$; instruiert-

te Kontrollprobanden: $M = 64.7$) noch signifikante Häufigkeitsunterschiede bei dichotomer Auswertung der SN-S. Dies könnte jedoch auch auf die geringe Stichprobengröße zurückzuführen sein, denn auffällige SN-S-Werte wurden von 43% der Begutachtungsprobanden, 50% der instruierten Kontrollprobanden, aber nur 25% der Kontrollprobanden erzielt. Zwischen Supernormalität und Intelligenz bestand kein signifikanter Zusammenhang. In der offen antwortenden Kontrollgruppe zeigte sich kein Zusammenhang zwischen Dissimulationsneigung und psychopathischen Persönlichkeitsmerkmalen. In den Stichproben der instruierten Kontrollprobanden und der Begutachtungsprobanden korrelierten dagegen PPI-R- und SN-S-Werte hoch miteinander ($r = .73$ bzw. $r = .67$), allerdings wie auch in den Studien von Cima et al. (2008; 2009) in der der Hypothese entgegengesetzten Richtung. Kein Proband mit hohen Psychopathiewerten antwortete supernormal. Bei der Diskussion der Ergebnisse geht der Autor zwar auf die geringe Stichprobengröße und damit die fragliche Repräsentativität seiner Befunde ein. Wie auch Cima et al. (2008; 2009) diskutiert er aber nicht die Möglichkeit, dass Probanden mit supernormalem Antwortverhalten die mit dem PPI erfassten, wenig sozial erwünschten, psychopathischen Persönlichkeitsmerkmale wie Kaltherzigkeit oder Egoismus verleugneten und daher in beiden Verfahren übereinstimmend niedrige Werte erzielten. Denn in beiden Studien wurde nicht die Kontrollskala des PPI, die manipulatives Antwortverhalten erfasst, verwendet, um Probanden mit unaufrichtigem Antwortverhalten im PPI aus der Analyse auszuschließen. Damit ist eine zuverlässige Beurteilung des Zusammenhangs nicht möglich, denn die PPI-Werte werden durch Antwortverzerrungen deutlich beeinflusst. Edens et al. (2001) konnten anhand einer studentischen Stichprobe zeigen, dass die PPI-Werte in einer Fake-good-Bedingung gegenüber dem Ausfüllen mit Standardinstruktion signifikant abnehmen. Dabei war die Differenz der Werte bei den Personen mit einer starken Ausprägung psychopathischer Merkmale deutlich größer. Eine Re-Analyse der Daten (Edens, 2004) zeigte jedoch, dass Impression Management nur auf den Skalen, die die sozial unerwünschten bzw. eher sanktionierten Eigenschaften erfassen (Schuldexternalisierung, Kaltherzigkeit, machiavellistischer Egoismus, sorglose Planlosigkeit), zu einer signifikanten Reduktion der Werte führte, jedoch nicht auf den Skalen zu eher unproblematischen Eigenschaften (sozialer Einfluss, Furchtlosigkeit, Stressimmunität).

Um zu klären, ob zwischen Supernormalität und Psychopathie tatsächlich ein negativer Zusammenhang besteht, wäre also eine Replikationsstudie vor allem mit Einsatz eines Fremdratings statt eines Selbstbeurteilungsverfahrens zur Erfassung psychopathischer Merkmale notwendig.

Von Zimmermann (2012) wurde ausgehend von der Studie von Wise (2002) der Zusammenhang zwischen dem Vorliegen einer Persönlichkeitsakzentuierung oder -störung und Supernormalität näher untersucht. Außerdem wurde der Einfluss des Begutachtungskontextes (Frage der Schuldfähigkeit vs. Frage der Rückfallprognose) noch einmal an einem größeren Probandenkollektiv analysiert. Darüber hinaus wurde ausgehend von den Ergebnissen von Cima et al. (2003c) die Hypothese geprüft, dass Probanden mit supernormalem Antwortverhalten bei der Deliktschilderung mehr kognitive Verzerrungen, wie Ablehnung der Verantwortung für die Tat, Verleugnung eines entstandenen Schadens etc., zeigen als Probanden ohne positive Antwortverzerrungen. Analysiert wurden die Akten von 62 Probanden (94% männlich), die zwischen Oktober 2010 und Mai 2012 in der Abteilung für Forensische Psychiatrie zur Frage der Schuldfähigkeit ($n = 31$) oder zur Frage der Rückfallprognose ($n = 31$) begutachtet worden waren. Die von Bantel (2010) bereits untersuchten Daten waren auch in diesem Kollektiv enthalten. Neben den SN-S-Ergebnissen wurden der Akte die klinischen Diagnosen gemäß ICD-10 (Dilling, Mombour, Schmidt & Schulte-Markwort, 2011) des psychiatrischen Sachverständigen sowie die Angaben zur Persönlichkeit des Probanden entnommen, die Aufschluss über das Vorliegen einer Persönlichkeitsakzentuierung oder -störung gaben. Anhand der im Gutachten zitierten Angaben des Probanden zum verfahrensgegenständlichen Delikt wurde das Vorliegen spezifischer kognitiver Verzerrungen anhand eines zuvor entwickelten Schemas kodiert. Probanden, die das Delikt vollständig leugneten ($n = 7$), wurden aus dieser Analyse ausgeschlossen. Die Hypothese, dass Prognoseprobanden mehr zu Dissimulation neigen als Schuldfähigkeitsprobanden, konnte nur teilweise bestätigt werden. Prognoseprobanden ($M = 59.3$) erzielten in der SN-S einen signifikant niedrigeren Mittelwert als Schuldfähigkeitsprobanden ($M = 70.5$). Auch auf den Skalen Affektive Symptome und Zwangssymptome unterschieden sich die beiden Stichproben signifikant in der erwarteten Richtung voneinander. Die dichotome Auswertung der SN-S erbrachte aber, trotz rein deskriptiv hypothesenkonformer Ergebnisse, keine signifikanten Unterschiede: 42% der Schuldfähigkeitsprobanden und 55% der Prognoseprobanden erzielten auffällige Werte. Hinsichtlich des Vorliegens negativer Antwortverzerrungen zeigten sich weder dimensional noch kategorial signifikante Gruppenunterschiede. Von den 44 Probanden mit Persönlichkeitsakzentuierung oder -störung zeigten 41% keine Antwortverzerrungen, 46% Supernormalität und 14% negative Antwortverzerrungen. Von den Probanden ohne klinische Persönlichkeitsauffälligkeiten zeigten 39% keine Antwortverzerrungen, 56% Supernormalität und 6% negative Antwortverzerrungen. Der Unterschied zwischen den beiden Gruppen war nicht signifikant. Auch die dimensio-

nale Auswertung der SN-S erbrachte keine signifikanten Gruppenunterschiede. Entsprechend der Hypothese zeigten aber Probanden mit supernormalem Antwortverhalten signifikant mehr Verzerrungen bei der Deliktschilderung als Probanden ohne Dissimulationsneigung.

Anhand der Daten der Eingangsdiagnostik einer psychotherapeutischen Fachambulanz für aus der Haft entlassene Sexualstraftäter mit Therapieweisung wurden die Auswirkung von supernormalem Antwortverhalten auf die Ergebnisse relevanter testpsychologischer Verfahren untersucht (Nedopil, Tippelt & Stübner, 2013). Es zeigte sich zum einen ein moderater bis starker Zusammenhang zwischen dem Ausmaß der Dissimulationsneigung und selbstberichteten deliktrelevanten kognitiven Verzerrungen bei Kindesmissbrauchern und Vergewaltigern. Je stärker die Probanden in der SN-S Beschwerden minimierten, desto geringer war das Ausmaß an kognitiven Verzerrungen in deliktspezifischen Fragebögen zu deliktfördernden Ansichten (Skala zur Erfassung kognitiver Verzerrungen bei Missbrauchern [Rambow, Elsner, Feelgood & Hoyer, 2008]: $N = 27$, $r = .71$; Skala zur Messung der Akzeptanz moderner Mythen über sexuelle Aggression [Gerger, Kley, Böhner & Siebler, 2007]: $N = 9$, $r = .47$) bzw. emotionaler Kongruenz mit Kindern (Skala zur Erfassung der emotionalen und kognitiven Überidentifikation mit der Kinderwelt [Wilson, 1999]: $N = 29$, $r = .64$).

Zum anderen wurde anhand dieser Stichprobe der Zusammenhang zwischen supernormalem Antwortverhalten und selbstberichteter Persönlichkeitspathologie näher untersucht, da erste Analysen gezeigt hatten, dass die Ergebnisse entsprechender Selbstbeurteilungsskalen nur bei wenigen Klienten der Fachambulanz mit der von den Therapeuten gestellten Diagnose einer Persönlichkeitsstörung übereinstimmten (Nedopil, Tippelt, Stübner, 2012). Der SN-S-Gesamtwert korrelierte signifikant moderat ($N = 53$, $r = .57$) mit der Anzahl unterdurchschnittlicher Skalenwerte in einem Inventar zur Erfassung von Persönlichkeitsakzentuierungen bzw. -störungen (Nedopil et al., 2013). Die Annahme, dass eine Vielzahl von unterdurchschnittlichen Skalen in einem solchen Inventar durch eine Neigung zur Minimierung von problematischen Persönlichkeitszügen mitbedingt ist, konnte somit gestützt werden.

Zwar liegen bereits einige Ergebnisse aus verschiedenen Untersuchungen mit der SN-S vor, eine umfassende Validierung, insbesondere der autorisierten deutschen Übersetzung, erfolgte bislang jedoch nicht. Zudem war der Stichprobenumfang bei Untersuchungen mit Begutachtungsprobanden zu gering, um Schlussfolgerungen verallgemeinern zu können und um differenzierte Analysen zu relevanten Fragestellungen, wie der Häufigkeit positi-

ver Antwortverzerrungen in Abhängigkeit vom Begutachtungskontext oder der Auswirkung von supernormalem Antworten auf testpsychologische Befunde, durchzuführen.

3 Fragestellungen

Ziel der vorliegenden Arbeit ist zunächst eine umfassende Überprüfung der überarbeiteten Fassung der SN-S durch die Bestimmung der wesentlichen Gütekriterien. Neben der Reliabilität soll die diagnostische Validität des Instruments bei der Kategorisierung des Antwortverhaltens bestimmt werden. Außerdem sollen im Rahmen der Konstruktvalidierung die Faktorenstruktur der SN-S und die Übereinstimmung mit anderen Skalen zur Erfassung von Antwortverzerrungen überprüft werden. Im Einzelnen sollen folgende Fragen beantwortet werden:

1. Wie hoch ist die Reliabilität des SN-S Gesamtwerts und ihrer Subskalen?
2. Lässt sich die Faktorenstruktur der Originalversion replizieren bzw. welche Subskalen können extrahiert werden?
3. Welche Cut-off-Werte sind bei SN-S-Gesamtwert und den einzelnen Subskalen für eine Kategorisierung des Antwortverhaltens heranzuziehen?
4. Wie zuverlässig ermöglicht die SN-S eine Einstufung des Antwortverhaltens als unauffällig oder supernormal? (diagnostische Validität)
5. Wie groß sind dimensionale und kategoriale Übereinstimmung mit anderen Kontrollskalen zur Erfassung von positiven Antwortverzerrungen (konvergente Validität)?
6. Erfasst die SN-S auch negative Antwortverzerrungen i.S. von Aggravation und Simulation?

In einem zweiten Schritt sollen die Häufigkeit von supernormalem Antwortverhalten bei forensischen Begutachtungen sowie dessen Auswirkungen auf relevante testpsychologische Befunde untersucht werden. Hier steht vor allem die praktische Relevanz der Anwendung der SN-S im Vordergrund, die Befunde können aber auch Sinne einer Überprüfung sowohl der Kriteriumsvalidität als auch der Konstruktvalidität (Hartig, Frey & Jude, 2012) interpretiert werden. Ausgehend von den bisherigen Studienergebnissen zur SN-S und den Auswirkungen von Antwortverzerrungen sollen folgende Hypothesen geprüft werden:

1. Das Auftreten positiver Antwortverzerrungen hängt von der Fragestellung der Begutachtung ab. Supernormalität tritt häufiger auf, wenn Probanden sich von der Begutach-

tung eine Entlassung aus der Haft bzw. dem Maßregelvollzug oder einen Nachweis ihrer Dienst- bzw. Geschäftsfähigkeit erhoffen, als bei Begutachtungen zur Frage der Schuldfähigkeit, bei denen Probanden ggf. auf eine Strafmilderung aufgrund psychischer Störung hoffen. Am seltensten tritt supernormales Antwortverhalten bei Probanden auf, die Rentenansprüche geltend gemacht oder Schadensersatzforderungen gestellt haben.

2. Probanden mit supernormalem Antwortverhalten berichten in Selbstbeurteilungsverfahren eine niedrigere Aggressionsbereitschaft und es zeigen sich häufiger Diskrepanzen zwischen selbstberichteter und mithilfe projektiver Verfahren erhobener Aggressionsbereitschaft als bei offen antwortenden Probanden.
3. Probanden mit supernormalem Antwortverhalten erzielen häufiger unterdurchschnittliche Skalenwerte in Inventaren zur Erfassung von Persönlichkeitsstilen und -störungen als offen antwortende Probanden.
4. Bei Probanden mit supernormalem Antwortverhalten stimmen Selbstbericht und klinische Einschätzung hinsichtlich des Vorliegens einer Persönlichkeitsstörung seltener überein als bei offen antwortenden Probanden.

4 Methoden

4.1 Die Stichprobe

Die Gesamtstichprobe setzt sich zusammen aus Kontrollprobanden, deren Daten online erhoben wurden, und Probanden, die in der Abteilung für Forensische Psychiatrie psychiatrisch und psychologisch begutachtet wurden. Zudem wurden die Daten von Bantel (2011) zur Kreuzvalidierung der mithilfe der Faktorenanalyse extrahierten Faktoren herangezogen.

4.1.1 Rekrutierung und Zusammensetzung der Kontrollstichprobe

Die Daten der Kontrollprobanden wurden mithilfe eines Onlinefragebogens erhoben, zum einen aus ökonomischen Gründen, vor allem aber, um einen möglichst hohen Grad an Anonymität zu gewährleisten und damit die Wahrscheinlichkeit positiver Antwortverzerrungen zu reduzieren. Aus diesem Grund wurden auch die IP-Adressen der Teilnehmer nicht erfasst. Der Fragebogen wurde mit dem Softwarepaket SoSciSurvey (Leiner, 2012) erstellt und umfasste eine kurze Erläuterung des Befragungsprojekts, in der die praktische

Relevanz des Forschungsvorhabens betont wurde, Fragen zu soziodemographischen Daten, die SN-S und die Kontrollskala Unaufrichtige Beantwortung des PPI-R (s. Kap. 4.3.2 Erhebungsinstrumente). Der Link zu dem Fragebogen wurde von den Mitarbeitern der Abteilung für Forensische Psychiatrie der LMU an Personen aus dem beruflichen und privaten Umfeld in einer Email verschickt, in der kurz auf Hintergrund und Ziele des Befragungsprojekts eingegangen und um Teilnahme an der Befragung sowie Weiterleitung des Links an andere Personen gebeten wurde. Aufgrund dieser Vorgehensweise konnte nicht nachvollzogen werden, wie viele Personen insgesamt den Link erhalten hatten.

Die Probanden wurden randomisiert entweder der Gruppe der „normalen“ Kontrollprobanden (KG), die möglichst offen die Fragen beantworten sollte, oder der Gruppe der „instruierten“ Kontrollprobanden (Fake-KG), die sich in einem möglichst günstigen Licht präsentieren, also supernormal antworten sollte, zugewiesen. Die Teilnehmer der KG erhielten lediglich die für den Fragebogen vorgesehenen Instruktionen zur Bearbeitung. Die Teilnehmer der Fake-KG erhielten vor Bearbeitung der Fragebögen eine Instruktion, die zu positiven Antwortverzerrungen führen sollte (s. Kap. 4.3.1).

Der Fragebogen war von 20. Oktober 2011 bis 05. März 2012 online freigeschaltet. Da die IP-Adressen der Teilnehmer nicht erfasst worden waren, war keine Unterscheidung möglich zwischen Personen, die den Fragebogen ansahen (sog. „Klicker“) und sich gänzlich gegen eine Teilnahme entschieden, und denen, die die Befragung dann auf einen späteren Zeitpunkt verschoben. Eine Analyse der Drop-outs unter Einbeziehung aller „Klicks“ war demnach nicht sinnvoll.

Insgesamt nahmen 290 Personen an der Befragung teil, d.h., dass sie mindestens ein Item beantworteten. Von ihnen waren 154 der KG und 136 der Fake-KG zugewiesen worden. Vorzeitig brachen insgesamt brachen 27 (9%) die Teilnahme ab, 17 Probanden der Fake-KG (13%) und 10 der KG (6%). Die Mehrheit der Abbrecher der Fake-KG ($n = 14$) beendete die Teilnahme nach Lesen der Fake-Instruktion (Seite 3).

Es verblieben 263 Probanden, 119 in der Fake-KG (88%) und 144 in der KG (93%).

Eine erste Datenanalyse zeigte, dass die Instruktion bei 49 Probanden der Fake-KG (41%) nicht wirksam war, weil sie in beiden Kontrollskalen unauffällig abschnitten, was gegen das Vorliegen positiver Antwortverzerrungen sprach. Dagegen gab es in der KG 12 Probanden (8%), die in beiden Kontrollskalen auffällige Werte erzielten, so dass bei ihnen vom Vorliegen von Antwortverzerrungen auszugehen war. Diese Probanden wurden von der Datenanalyse ausgeschlossen, so dass 70 Probanden in der Fake-KG und 131 Proban-

den in der KG verblieben. Es wurden also Daten von 69% der 290 Personen, die an der Befragung teilgenommen hatten, in die Auswertung einbezogen.

4.1.2 Zusammensetzung der Stichprobe der Begutachtungsprobanden

In die Stichprobe aufgenommen wurden Daten aller Probanden, die zwischen Oktober 2010 und Juni 2012 in der Abteilung für Forensische Psychiatrie der LMU psychiatrisch begutachtet wurden und bei denen eine testpsychologische Untersuchung des Antwortverhaltens mit der SN-S erfolgte. Eine Bearbeitung von Fragebögen setzt ein Mindestmaß an kognitiven und sprachlichen Fähigkeiten voraus, so dass bei einem IQ unter 80 oder mangelnden Sprachkenntnissen auf ihren Einsatz bei Begutachtungen verzichtet wird. Die Untersuchung des Antwortverhaltens war in den meisten Fällen routinemäßiger Bestandteil einer ausführlicheren (test-)psychologischen Begutachtung. Einigen Probanden wurden die Fragebögen aber lediglich im Rahmen der psychiatrischen Untersuchung vorgelegt und von der Autorin ausgewertet. Insgesamt wurden Daten von 108 Personen in die Analyse eingeschlossen, davon wurden 67 durch die Autorin (62%) und 32 durch andere psychologische Mitarbeiter der Abteilung (30%) testpsychologisch untersucht, die übrigen 10 Probanden (9%) bearbeiteten die Fragebögen im Rahmen der psychiatrischen Begutachtung. Die Probanden wurden entsprechend der Fragestellung des psychiatrischen Hauptgutachtens und den damit verbundenen unterschiedlichen Zielsetzungen der Probanden in vier Gruppen eingeteilt:

- „Prognose“ ($n = 43$): Dieser Gruppe wurden alle Probanden zugeteilt, die sich zum Zeitpunkt der Begutachtung in Haft ($n = 10$), im Maßregelvollzug gem. § 63 ($n = 22$) oder § 64 StGB ($n = 2$) oder in der Sicherungsverwahrung gem. § 66 StGB ($n = 9$) befanden und im Auftrag des Gerichts oder der JVA zur Frage der Rückfallprognose im Falle von Lockerungen oder einer Entlassung begutachtet wurden.
- „Schuldfähigkeit“ ($n = 34$): Hier wurden alle Probanden subsumiert, die im Auftrag der Staatsanwaltschaft oder des Gerichts zur Frage der Schuldfähigkeit gem. § 20, 21 StGB und ggf. zur Notwendigkeit einer Unterbringung im Maßregelvollzug gem. §§ 63 oder 64 StGB ($n = 26$) bzw. in der Sicherungsverwahrung gem. § 66 StGB ($n = 1$) und/oder zur Frage der Reife gem. § 105 JGG ($n = 7$) begutachtet wurden.
- „Zahlungsbegehren“ ($n = 23$): Alle Probanden, die im Auftrag eines Gerichts oder einer Versicherung begutachtet wurden, nachdem sie Ansprüche auf Zahlungen i.S. von Erwerbsminderungs- ($n = 9$), Berufsunfähigkeitsrente ($n = 8$), Schadensersatz und/oder

Schmerzensgeld ($n = 5$) oder Opferentschädigung ($n = 1$) geltend gemacht hatten, wurden dieser Gruppe zugewiesen.

- „Gesundheitsnachweis“ ($n = 8$): In diese Gruppe wurden Probanden eingeteilt, die im Auftrag ihres Dienstherrn begutachtet wurden, der Zweifel an der Dienstfähigkeit des Probanden hatte, die selbst jedoch ihre Arbeitsstelle behalten wollten ($n = 4$), außerdem Probanden, die zur Frage der Geschäftsfähigkeit begutachtet wurden ($n = 2$). Ferner wurde eine Probandin, die zur Frage der Erziehungsfähigkeit mit drohendem Entzug des Sorgerechts für ihre Kinder untersucht wurde, und ein Proband, der zur Frage der Einrichtung einer Betreuung begutachtet wurde, die gegen seinen Willen installiert werden sollte, hier subsumiert.

4.1.3 Beschreibung der Stichprobe

Die Gesamtstichprobe bestand aus 309 Probanden, die den oben beschriebenen sechs Untersuchungsgruppen zugeordnet waren. Die einzelnen Gruppen unterschieden sich erwartungsgemäß hinsichtlich verschiedener demographischer Merkmale, für die jedoch kein Zusammenhang mit dem Antwortverhalten zu erwarten war.

Die Substichproben unterschieden sich erwartungsgemäß signifikant hinsichtlich Alter, $F(5, 301) = 12.41, p < .001$, und Geschlecht, $\chi^2(5, N = 308) = 61.25, p < .001$. Die Probanden der Prognose- und der Gesundheitsnachweis-Gruppe waren älter als die der anderen Gruppen (s. Tabelle 1).

Tabelle 1

Alter der untersuchten Probanden in Jahren

Stichprobe	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Schuldfähigkeit	34	35.15	14.40	18	62
Prognose	43	49.56	12.20	21	74
Zahlungsbegehren	23	44.87	9.73	24	59
Gesundheitsnachweis	8	53.25	18.17	22	73
Kontrollgruppe	130	35.03	13.04	18	74
Fake-Kontrollgruppe	69	37.01	11.61	18	64

Auch hinsichtlich des Geschlechts waren die Stichproben nicht vergleichbar. In der KG- und der Fake-KG war der Anteil der Frauen erwartungsgemäß höher, während die Strafrechtsprobanden überwiegend bzw. ausschließlich männlich waren (s. Tabelle 2).

Tabelle 2

Demographische Merkmale der untersuchten Probanden (Häufigkeit, Prozentwerte in Klammern) (Fortsetzung auf S. 37)

	SF	RP	ZB	GN	KG	Fake-KG
Merkmal	(n = 34)	(n = 43)	(n = 23)	(n = 8)	(n = 131)	(n = 70)
Geschlecht ^a						
männlich	30 (88.2)	43 (100)	13 (56.5)	6 (75.0)	56 (43.1)	32 (45.7)
weiblich	4 (11.8)	-	10 (43.5)	2 (25.0)	74 (56.9)	38 (54.3)
kultureller Hintergrund ^b						
deutsch	23 (67.6)	40 (93.0)	22 (95.7)	7 (87.5)	129 (98.5)	65 (94.2)
nicht deutsch	11 (32.4)	3 (7.0)	1 (4.3)	1 (12.5)	2 (1.5)	4 (5.8)
Familienstand ^c						
ledig	25 (73.5)	20 (46.5)	10 (43.5)	3 (37.5)	79 (60.3)	34 (50.0)
verheiratet	-	5 (11.6)	7 (30.4)	1 (12.5)	42 (32.1)	31 (45.6)
getrennt/geschieden/ verwitwet	9 (26.5)	18 (41.9)	6 (26.1)	4 (50.0)	10 (7.6)	3 (4.4)

Fortsetzung Tabelle 2	SF (n = 34)	RP (n = 43)	ZB (n = 23)	GN (n = 8)	KG (n = 131)	Fake-KG (n = 70)
Schulabschluss^d						
keiner	9 (26.5)	7 (16.3)	-	-	-	-
Förderschulab- schluss	1 (2.9)	2 (4.7)	-	-	-	-
(Qualifizierender) Hauptschulabschluss	13 (38.2)	22 (51.2)	12 (52.2)	-	12 (9.4)	8 (11.4)
Mittlere Reife	6 (17.6)	9 (20.9)	3 (13.0)	5 (62.5)	23 (18.0)	19 (27.1)
(Fach-)Abitur	5 (14.7)	3 (7.0)	8 (34.8)	3 (37.5)	93 (72.7)	43 (61.4)
Berufsausbildung^e						
keine	15 (44.1)	12 (27.9)	2 (8.7)	1 (12.5)	7 (5.4)	2 (2.9)
Ausbildung/Lehre	12 (35.3)	31 (72.1)	17 (73.9)	3 (37.5)	43 (33.1)	25 (36.2)
Fach-/Meisterschule	2 (5.9)	-	-	1 (12.5)	7 (5.4)	6 (8.7)
(Fach-)Hochschul- abschluss	3 (8.8)	-	4 (17.4)	3 (37.5)	52 (40.0)	29 (42.0)
aktuell in Ausbildung	2 (5.9)	-	-	-	21 (16.2)	7 (10.1)

Anmerkung: ^{a, b, c, d, e} $p < .001$

Wie in Tabelle 2 zu sehen, unterschieden sich die Gruppen auch hinsichtlich weiterer demographischer Merkmale. Die Probanden der beiden Kontrollgruppen waren häufiger verheiratet und wiesen einen höheren schulischen und beruflichen Bildungsstand auf als die Strafrechtsprobanden. In der Gruppe der Schuldfähigkeitsprobanden waren mehr Personen mit einem nicht-deutschen kulturellen Hintergrund vertreten.

Die Begutachtungsprobanden unterschieden sich signifikant hinsichtlich der Anzahl psychiatrischer Vorgutachten, $F(3, 104) = 12.04, p < .001$. Die meisten Vorgutachten hatten erwartungsgemäß die Probanden der Gruppen Prognose ($M = 3.56, SD = 2.52$) und Zahlungsbegehren ($M = 2.04, SD = 2.92$), während die Gesundheitsnachweis- ($M = 1.3, SD = 1.75$) und Schuldfähigkeitsprobanden ($M = 0.53, SD = 1.19$) seltener psychiatrisch begutachtet worden waren.

Bei 26 Probanden (24%) war vom psychiatrischen Gutachter keine psychische Störung i.S. der ICD-10 festgestellt worden. Die Hauptdiagnosen (F-Kategorie) der anderen Probanden können Tabelle 3 entnommen werden. Eine Persönlichkeitsstörung war bei 29 Probanden (27%) diagnostiziert worden, dabei handelte es sich ausschließlich um Strafrechtsprobanden. Am häufigsten traten die dissoziale (41%) und die kombinierte Persönlichkeitsstörung (24%) auf (s. Tabelle 3).

Tabelle 3

Vom psychiatrischen Gutachter gestellte Hauptdiagnosen (ICD-10 F-Kategorien) und diagnostizierte Persönlichkeitsstörungen (PS) (Häufigkeit, Prozentwerte in Klammern) (Fortsetzung auf S. 39)

	SF (<i>n</i> = 34)	RP (<i>n</i> = 43)	ZB (<i>n</i> = 23)	GN (<i>n</i> = 8)	Gesamt (<i>N</i> = 108)
keine ICD-Diagnose	6 (17.6)	11 (25.6)	5 (21.7)	4 (50.0)	26 (24.1)
F0	1 (2.9)	1 (2.3)	-	1 (12.5)	3 (2.8)
F1	10 (29.4)	6 (14.0)	-	1 (12.5)	17 (15.7)
F2	-	6 (14.0)	-	1 (12.5)	7 (6.5)
F3	1 (2.9)	-	9 (39.1)	-	10 (9.3)
F4	2 (5.9)	-	9 (39.1)	1 (12.5)	12 (11.1)
F6	12 (35.3)	17 (39.5)	-	-	29 (26.9)
F9	2 (5.9)	2 (4.7)	-	-	4 (3.7)

Fortsetzung Tabelle 3	SF (n = 34)	RP (n = 43)	ZB (n = 23)	GN (n = 8)	Gesamt (N = 108)
keine PS	23 (21.3)	25 (58.1)	23 (100)	8 (100)	79 (73.1)
dissoziale PS (F60.2)	3 (8.8)	9 (20.9)	-	-	12 (11.1)
emotional-instabile PS (F60.3)	3 (8.8)	1 (2.3)	-	-	4 (3.7)
narzisstische PS (F60.8)	-	1 (2.3)	-	-	1 (0.9)
PS nicht näher bezeichnet (F60.9)	-	1 (2.3)	-	-	1 (0.9)
kombinierte PS (F61)	3 (8.8)	4 (9.3)	-	-	7 (6.5)
organische PS (F07.0)	1 (2.9)	2 (4.7)	-	-	3 (2.8)
PS durch psychotrope Substanzen (F19.7)	1 (2.9)	-	-	-	1 (0.9)

Die Indexdelikte der Strafrechtsprobanden können Tabelle 4 entnommen werden. Die Prognoseprobanden hatten häufiger schwere Taten (Tötungs- und Sexualdelikte) begangen als die Schuldfähigkeitsprobanden ($p < .01$).

Tabelle 4

Indexdelikte der Strafrechtsprobanden (Häufigkeit, Prozentwerte in Klammern)

Indexdelikt	SF (n = 34)	RP (n = 43)	Gesamt (N = 77)
Mord/Totschlag	3 (8.8)	15 (34.9)	18 (23.4)
Körperverletzung	9 (26.5)	6 (14.0)	15 (19.5)
Raub	6 (17.6)	3 (7.0)	9 (11.7)
Sexualdelikt	3 (8.8)	17 (39.5)	20 (26.0)
Diebstahl	1 (2.9)	-	1 (1.3)
Betrug	3 (8.8)	-	3 (4.0)
BtMG-Verstoß	3 (8.8)	1 (2.3)	4 (5.2)
Sonstiges	6 (17.6)	1 (2.3)	7 (9.1)

4.2 Stichprobe zur Kreuzvalidierung (Bantel, 2011)

Die Daten von Bantel (2011; s. Kap. 2.5) wurden zur Kreuzvalidierung der Subskalen, die durch eine Faktorenanalyse extrahiert wurden, herangezogen.

Von diesen Probanden waren neben ausgefüllten SN-S- und PPI-R-Bögen lediglich Angaben zu Alter und Geschlecht verfügbar.

Zunächst wurde nach demselben Procedere wie in Kapitel 4.1.1 beschrieben eine Stichprobenkorrektur durchgeführt. Dies führte zum Ausschluss von 8 Probanden aus der Fake-KG. Die Stichprobe, im folgenden KV (Kreuzvalidierung) genannt, bestand dann aus 28 Probanden in der KG und 20 Probanden in der Fake-KG.

Angaben zu Alter und Geschlecht können Tabelle 5 entnommen werden. Die beiden Gruppen unterschieden sich diesbezüglich nicht signifikant. Sie unterschieden sich hinsichtlich des Geschlechts außerdem nicht signifikant von den beiden Stichproben der Onlinebefragung, allerdings waren die Probanden der KV im Schnitt etwas älter, $F(3, 242) = 32.98, p < .001$.

Tabelle 5

Alter in Jahren und Geschlecht (Häufigkeit, Prozentwerte in Klammern) der Probanden von Bantel (2011) nach Stichprobenkorrektur

Alter	<i>M (SD)</i>	<i>Min - Max</i>
Gesamt ($N = 48$)	54.83 (11.57)	21 - 72
KG ($n = 28$)	55.07 (13.55)	21 - 72
Fake-KG ($n = 20$)	53.74 (8.56)	37 - 63
Geschlecht	männlich	weiblich
Gesamt ($N = 48$)	21 (43.8)	27 (56.3)
KG ($n = 28$)	11 (39.3)	17 (60.7)
Fake-KG ($n = 20$)	10 (50.0)	10 (50.0)

4.3 Untersuchungsdesign und Erhebungsinstrumente

Zur Beantwortung der in Kapitel 3 aufgeführten Fragestellungen wurde das Antwortverhalten von Kontrollprobanden aus der Allgemeinbevölkerung, von denen ein Teil instruiert wurde, sich in möglichst günstigem Licht zu präsentieren, und vier verschiedenen Gruppen

von Begutachtungsprobanden mithilfe der SN-S und weiterer Instrumente überprüft. Während der Erhebungsablauf für die Kontrollgruppen standardisiert war, kamen bei den Begutachtungen unterschiedliche Verfahren zum Einsatz, die in keiner festgelegten Reihenfolge vorgegeben worden waren. Die Befunde wurden nicht speziell für die vorliegende Studie erhoben, daher unterscheiden sich die Fallzahlen bei den einzelnen Auswertungen. So wurde z.B. die Aggressionsbereitschaft lediglich bei den Probanden untersucht, die Gewaltdelikte (Körperverletzung, Tötungs- oder gewalttätige Sexualdelikte) begangen hatten, bei denen also die Erfassung relevant für die gutachtliche Beurteilung war. Inventare zur Erfassung von Persönlichkeitsstilen und -störungen kamen bei den Begutachtungen zum Einsatz, wenn sich aus den Akten Hinweise auf das Vorliegen klinisch relevanter Persönlichkeitsmerkmale bzw. einer Persönlichkeitsstörung ergaben. Im Folgenden wird für jedes der Verfahren beschrieben, zu welchem Zweck es bei der Begutachtung eingesetzt wurde bzw. von wie vielen Probanden das Instrument jeweils vorlag.

Die Bildung der beiden Kontrollgruppen KG und Fake-KG diente der Berechnung der diagnostischen Validität. Die Ergebnisse anderer Instrumente bzw. Skalen zu positiven Antwortverzerrungen wurden zur Berechnung der konvergenten Validität herangezogen. Die Ergebnisse des Strukturierten Fragebogens Simulierter Symptome (SFSS; Cima et al., 2003b) dienten zur Überprüfung der Frage, ob die SN-S auch negative Antwortverzerrungen erfasst. Zum Zwecke der Konstruktvalidierung wurde die Aggressionsbereitschaft mithilfe eines Selbstbeurteilungsverfahrens und zweier projektiver Verfahren bestimmt. Ob laut Gutachter eine Persönlichkeitsstörung vorlag, wurde dem psychiatrischen Gutachten entnommen. Anhand verschiedener spezifischer Inventare zur Erfassung von Persönlichkeitsstilen und -störungen wurde geprüft, ob nach Selbsteinschätzung der Probanden eine Persönlichkeitsstörung vorlag. Außerdem wurde die Anzahl unterdurchschnittlicher Skalenwerte in diesen Inventaren erhoben.

4.3.1 Onlinefragebogen

Der Onlinefragebogen wurde mithilfe des Softwarepakets SoSciSurvey (Leiner, 2012) programmiert. Der erste Entwurf des Onlinefragebogens wurde von einem erfahrenen forensischen Psychiater und einer erfahrenen forensischen Psychologin in einem Pretest ausgefüllt und entsprechend ihrer Anmerkungen überarbeitet.

Auf der ersten Seite wurden die Teilnehmer über Ziele und Hintergründe der Studie aufgeklärt. Durch die Formulierung sollte zum einen die Bedeutsamkeit der Untersuchungsergebnisse betont, zum anderen die wahrgenommene Anonymität der Befragung gestärkt

werden. Auf der zweiten Seite wurden demographische Angaben erhoben (Alter, Geschlecht, Familienstand, Staatsangehörigkeit, Muttersprache, Schulabschluss, Berufsausbildung und Berufstätigkeit). Für die dritte Seite wurde ein Filter programmiert, der die Probanden randomisiert einer der beiden Gruppen (KG oder Fake-KG) zuwies. Probanden der Fake-Bedingung erhielten dann auf Seite 3 des Onlinefragebogens zunächst folgende spezifische Instruktion, durch deren Formulierung die Ähnlichkeit mit der Begutachtungssituation und die Motivation für positive Antwortverzerrungen durch Betonung der Bedeutsamkeit der Ergebnisse erhöht werden sollten:

„Stellen Sie sich vor, Sie befinden sich in einer **Begutachtungssituation** bei einem Psychiater bzw. Psychologen, deren Ergebnis über den weiteren Verlauf Ihrer Unterbringung in einer forensischen Einrichtung (Haft oder Maßregelvollzug) entscheidet. Lockerungen oder eine Entlassung werden umso wahrscheinlicher, je glaubhafter Sie deutlich machen können, dass Sie **nicht mehr gefährlich** sind. Behalten Sie dies im Hinterkopf und antworten Sie im Folgenden so, wie Sie glauben, dass es für Sie **am günstigsten** wäre.

Sollten Sie sich bei einer Antwort einmal nicht sicher sein, so kreuzen Sie bitte die am ehesten zutreffende Antwort an. **Bitte geben Sie immer eine Antwort**, da sonst im ungünstigsten Fall der komplette Fragebogen unbrauchbar wird.“

Dann folgte für beide Gruppen auf jeweils einer Seite die SN-S und die Skala Unaufrichtige Beantwortung des PPI-R (PPI-UA; s.u.). Für die SN-S wurde die kurze Instruktion der Originalversion beibehalten, die umfangreiche Instruktion des PPI-R wurde gekürzt:

„Bitte lesen Sie sich die folgenden 23 Aussagen genau durch und entscheiden Sie, wie sehr jede Aussage auf Sie zutrifft oder nicht, wie falsch oder richtig sie für Sie ist. Dann markieren Sie die zutreffendste Antwort. Benutzen Sie folgende Antwortskala:

falsch, eher falsch, eher richtig, richtig.

Selbst wenn Sie das Gefühl haben, dass eine Aussage für Sie weder falsch noch richtig ist, wählen Sie bitte die Antwort, die Sie am ehesten beschreibt.“

Auf der letzten Seite wurde für die Teilnahme gedankt. Der Onlinefragebogen bestand somit insgesamt für die KG aus fünf Seiten, für die Fake-KG aus sechs Seiten.

Um die Anzahl der Abbrecher zu reduzieren, wurde auf ein Erzwingen von Antworten bei einzelnen Items verzichtet. Da für beide Instrumente kein spezieller Umgang mit fehlenden

Werten, wie das Ersetzen durch den Mittelwert, vorgesehen ist, wurden fehlende Werte mit 0 kodiert. Die Anzahl fehlender Werte war für die einzelnen Items beider Instrumente gering (SN-S: 0-2 Werte; PPI-UA: 3-6 Werte).

4.3.2 Psychopathic Personality Inventory-Revised (PPI-R)

Das PPI-R (Alpers & Eisenbarth, 2008) ist ein Fragebogen zur Erfassung der dimensionalen Ausprägung psychopathischer Persönlichkeitsmerkmale. Es besteht aus 154 Aussagen, die auf einer vierstufigen Skala (von „falsch“ bis „richtig“) einzuschätzen sind. Die Items sind den Dimensionen Schuldexternalisierung, Rebellen Risikofreude, Stressimmunität, Sozialer Einfluss, Kältherzigkeit, Machiavellistischer Egoismus, Sorglose Planlosigkeit und Furchtlosigkeit zugeordnet. Aus diesen kann auch ein Gesamtwert für das Merkmal Psychopathie gebildet werden. Außerdem besteht die Möglichkeit, durch die Skala Unaufrichtige Beantwortung (PPI-UA) Antworttendenzen manipulativer Art zu überprüfen. Im Manual werden geschlechtsspezifische Normwerte zur Verfügung gestellt (T-Werte), die allerdings anhand einer rein studentischen Stichprobe gewonnen wurden.

Die UA-Skala besteht aus 23 Items, die zum einen sozial erwünschte Antworten, zum anderen systematisches Ankreuzen ohne Lesen der Iteminhalte erfassen. Aufgrund dieser Heterogenität überrascht die im Manual berichtete geringe interne Konsistenz (Cronbachs $\alpha = .65$) nicht. Für die anderen PPI-Skalen werden Werte zwischen $\alpha = .72$ und $\alpha = .88$ berichtet. In einer Studie von Eisenbarth & Alpers (2007) wiesen forensische Probanden im Vergleich zu männlichen Studenten signifikant höhere Werte im Gesamtscore und auf den Skalen Schuldexternalisierung, Rebellen Risikofreude und Unaufrichtige Beantwortung auf. Die PPI-UA-Werte korrelierten signifikant mit den Ergebnissen einer Skala zur sozialen Erwünschtheit ($r = .30$). Zumindest für die englischsprachige Originalversion konnten signifikante moderate Korrelationen mit der PCL-R gezeigt werden (Berardino, Meloy, Sherman & Jacobs, 2005; Poythress, Edens & Lilienfeld, 1998).

Das PPI-R wird bei der Begutachtung von Probanden mit dissozialen Persönlichkeitseigenschaften als Screening-Instrument eingesetzt. In der vorliegenden Arbeit wurde die UA-Skala für die Überprüfung der Konstruktvalidität herangezogen und daher in den Onlinefragebogen integriert. Zudem diente die Skala der Stichprobenkorrektur (s. Kap. 4.1.1). Von den Begutachtungsprobanden wurden neben den 23 Items Roh- und T-Wert der UA-Skala aus der Akte erhoben. Außerdem wurde aufgrund der Heterogenität der Skala ein Summenwert für die Items, die soziale Erwünschtheit messen, gebildet. Diese neue Skala (PPI-SE) wurde ebenfalls zur Überprüfung der Konstruktvalidität herangezogen.

Die UA-Skala lag von allen 129 KG- und 68 Fake-KG-Probanden vor. Das gesamte PPI-R war von 41 Begutachtungsprobanden (Schuldfähigkeit: $n = 19$; Prognose: $n = 22$) bearbeitet worden.

4.3.3 Persönlichkeits-Stil- und Störungsinventar (PSSI)

Das PSSI (Kuhl & Kazén, 2009) ist ein Selbstbeurteilungsinstrument zur Erfassung der relativen Ausprägung von Persönlichkeitsstilen. Diese gelten als nichtpathologische Entsprechungen der in DSM-IV und ICD-10 beschriebenen Persönlichkeitsstörungen. Der Fragebogen besteht aus 140 Items, die jeweils auf einer vierstufigen Skala einzuschätzen sind (von „trifft gar nicht zu“ bis „trifft ausgesprochen zu“). Je zehn Aussagen sind einer von insgesamt 14 Skalen zugeordnet. Die Summenwerte der einzelnen Skalen werden anhand geschlechts- und altersspezifischer Normtabellen in T-Werte transformiert. Überdurchschnittliche Werte (T-Wert > 60) können als Hinweis auf das Vorliegen einer Persönlichkeitsakzentuierung oder -störung gesehen werden. Eine Diagnosestellung allein aufgrund der Testbefunde ist jedoch nicht möglich. Dagegen kann eine Vielzahl niedriger Werte (T-Wert < 40) laut Handbuch eher für einen bedeckten Antwortstil sprechen bzw. dafür, dass die Person nicht willens oder in der Lage ist, über ihre Persönlichkeit Auskunft zu geben.

Die Reliabilität (Cronbachs α) der einzelnen PSSI-Skalen lag in der Konstruktionsstichprobe ($N = 1\,227$) zwischen $\alpha = .73$ und $\alpha = .85$, die Retest-Reliabilität zwischen $r_{tt} = .68$ und $r_{tt} = .83$. Das Verfahren ist testtheoretisch und konzeptuell persönlichkeitspsychologisch gut begründet und es liegt eine Reihe von Studien zu korrelativen Zusammenhängen zwischen den PSSI-Skalen und konstrukt-nahen Persönlichkeitsmerkmalen bzw. klinischen Merkmalen vor. Die Konkordanz der Selbsteinschätzung mit der klinischen Diagnose wurde bisher allerdings nicht untersucht. Das PSSI wird daher auch in der S2 Leitlinie für Persönlichkeitsstörungen lediglich als diagnostisches Hilfsmittel empfohlen (Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde, 2009).

Dieses Verfahren wird bei Begutachtungen bei Verdacht auf eine Persönlichkeitsstörung als Screening-Instrument eingesetzt. Außerdem kann es gerade bei Prognosebegutachtungen durch den Vergleich von Selbst- und Fremdeinschätzung Hinweise auf das Ausmaß der Störungseinsicht des Probanden geben. Ausgefüllte Bögen lagen von 24 Probanden vor (Schuldfähigkeit: $n = 8$, Prognose $n = 13$, Zahlungsbegehren: $n = 1$, Gesundheitsnachweis: $n = 2$). Aus der Akte wurden nicht alle Ergebnisse des Fragebogens, sondern lediglich die

Anzahl der Skalen, auf denen ein unterdurchschnittlicher Wert (T-Wert < 40) erzielt wurde, entnommen.

4.3.4 Inventar Klinischer Persönlichkeitsakzentuierungen (IKP)

Das IKP (Andresen, 2006) ist ein Selbstbeurteilungsfragebogen, das der dimensionalen Erfassung von Persönlichkeitsakzentuierungen nach DSM-IV und ICD-10 dient. Mit dem Grundinventar (IKP-G) können 11 offizielle Diagnoseeinheiten nach DSM-IV und ICD-10 erfasst werden. Das Ergänzungsmodul (IKP-Eg) erfasst auf 13 Skalen differentialdiagnostisch relevante Persönlichkeitsstile, Spektrums-Bereiche bzw. persönlichkeitsnahe Subsyndrome und traditionelle Persönlichkeitsstörungen nach Kurt Schneider, die nicht in den Diagnosesystemen enthalten sind. Dadurch wird eine differenzierte und umfassende Einschätzung von problematischen Persönlichkeitszügen und Interaktionsschwierigkeiten ermöglicht. Jede der insgesamt 24 Skalen umfasst zehn Items, die auf einer vierstufigen Skala zu beurteilen sind (von „völlig unzutreffend“ bis „völlig zutreffend“). Dabei handelt es sich ausschließlich um positiv, also in symptomatischer Richtung, gepolte Aussagen. Zwar gibt es pro Skala auch je zwei negativ gepolte Items, diese gehen aber nicht in die Skalenbildung ein. Die Skalenwerte können anhand einer Normtabelle in T-Werte transformiert werden. Allerdings stehen bisher lediglich Grobnormen zur Verfügung, die anhand einer zwar soziodemographisch heterogenen, aber nicht repräsentativen Normalstichprobe gewonnen wurden. Als Grenzwert für die Verdachtsdiagnose einer Persönlichkeitsstörung gilt ein T-Wert von 70 Punkten, wobei ab 65 Punkten schon eine starke Ausprägung der Persönlichkeitsakzentuierung vorliegt und die Wahrscheinlichkeit für eine Persönlichkeitsstörung erhöht ist. Eine Diagnosestellung allein aufgrund der Testbefunde ist jedoch ebenso wie mit dem PSSI nicht möglich.

Im Handbuch werden zufriedenstellende Gütekriterien berichtet. Die interne Konsistenz der einzelnen Skalen liegt zwischen $\alpha = .73$ und $\alpha = .93$, Retest-Reliabilitäten werden nur für die ursprüngliche Langfassung des Fragebogens mit 18 Items pro Skala berichtet und liegen zwischen $r_{tt} = .72$ und $r_{tt} = .85$. Es werden außerdem paarweise Korrelationen zwischen IKP- und entsprechenden PSSI-Skalen zwischen $r = .48$ und $r = .84$ berichtet. Schröder (2004) untersuchte den Zusammenhang zwischen den Ergebnissen (dimensionale Scores) eines diagnostischen Interviews für Persönlichkeitsstörungen (IPDE, Mombour et al., 1996) und den IKP-G-Ergebnissen an 41 stationär-psychiatrischen Patienten mit Cluster-C-Persönlichkeitsstörungen. Für sechs von neun Diagnosen fanden sich signifikante Korrelation mit Koeffizienten zwischen $r = .39$ (Impulsive Persönlichkeitsstörung) und $r =$

.52 (Borderline-Persönlichkeitsstörung). Keine signifikanten Zusammenhänge mit der klinischen Diagnose fanden sich dagegen auf den Skalen Paranoide ($r = .29$), Dissoziale ($r = -.36$) und Histrionische Persönlichkeitsstörung ($r = .25$). In einer Studie von Henkel (2005) wurde die Konkordanz zwischen der klinischen Diagnose Borderline-Persönlichkeitsstörung und dem Ergebnis der entsprechenden IKP-G-Skala an 71 stationär-psychiatrischen Patienten untersucht. Mit einem Kappa-Wert von .48 wurde eine moderate Übereinstimmung erzielt. Dabei war vor allem die Anzahl der falsch Positiven im IKP-G hoch (16/30), die Rate der falsch Negativen dagegen eher gering (3/41).

Das IKP wird bei Begutachtungen ähnlich wie das PSSI eingesetzt. Da das Grundinventar im Gegensatz zum PSSI auch eine Skala für den impulsiven Typus der emotional-instabilen Persönlichkeitsstörung enthält, wird es bei Probanden mit impulsiv-aggressiven Delikten häufiger angewendet. Außerdem wird oft das IKP-Eg mit IKP-G oder PSSI vorgegeben, wenn weitere Persönlichkeitsauffälligkeiten und Interaktionsschwierigkeiten erfasst werden sollen, wie z.B. Willensschwäche bei Probanden mit Substanzproblematik. Ausgefüllte IKP-G-Bögen lagen von 27 Probanden (Schuldfähigkeit: $n = 12$, Prognose $n = 14$) und IKP-Eg-Bögen von 24 Probanden (Schuldfähigkeit: $n = 3$, Prognose $n = 21$) vor. Auch hier wurde jeweils nur die Anzahl der Skalen, auf denen ein unterdurchschnittlicher Wert (T-Wert < 40) erzielt wurde, aus der Akte entnommen.

4.3.5 Fragebogen zur Erfassung von Aggressivitätsfaktoren (FAF)

Der FAF (Hampel & Selg, 1998) ist ein Selbstbeurteilungsinstrument zur Erfassung der Bereitschaft zu aggressiven Verhaltensweisen. Er besteht aus 77 Items, die auf einer zweistufigen Skala im Ja-Nein-Format beurteilt werden müssen. Daraus werden die Skalen Spontane Aggression, Reaktive Aggression, Erregbarkeit, Selbstaggression, Aggressionshemmung sowie die Skala Offenheit (FAF-Off; s. Anhang A4) gebildet, die die Bereitschaft erfasst, kleinere Schwächen einzugestehen. Laut Handbuch werden allerdings auch von Personen mit hohen Hemmungen niedrige Offenheitswerte erreicht. Aus den ersten drei Skalen kann ein Gesamtwert „ Σ Aggressivität“ berechnet werden, der die nach außen gerichtete Aggressionsbereitschaft erfasst. Die Skalenrohwerte (Summenwerte) können anhand von alters- und geschlechtsspezifischen Tabellen in Prozentrang-, Stanine- und T-Werte transformiert werden. Der FAF besitzt zufriedenstellende Gütekriterien. Die interne Konsistenz (Cronbachs α) der Skalen und des Gesamtwerts liegt zwischen $\alpha = .61$ (FAF-Off) und $\alpha = .85$. Erwachsene inhaftierte Straftäter erreichen signifikant höhere Werte als Kontrollprobanden auf den ersten vier Skalen sowie einen signifikant höheren Gesamtwert.

Zwischen Straftätergruppen mit verschiedenen Indexdelikten (Eigentums-, Körperverletzungs-, Sexual- und sonstige Delikte) fanden sich allerdings keine signifikanten Unterschiede. Kury (1987) konnte zeigen, dass Strafgefangene, die aufgrund der Testresultate (negative) soziale Konsequenzen fürchteten, mehr zu Verfälschungstendenzen neigten und aggressive Verhaltensweisen mehr verneinten als Strafgefangene, die den Fragebogen anonym ausfüllten. Trotz allem gehört der FAF zu den bei forensischen Begutachtungen am häufigsten eingesetzten Verfahren (z.B. Kunzl, 2011). Denn die Gegenüberstellung von Fremd- und Selbsteinschätzung ermöglicht Aussagen über die Störungseinsicht von Probanden. Aufgrund der Anfälligkeit für Verzerrungstendenzen wird der FAF aber häufiger bei Schuldfähigkeits- als bei Prognosebegutachtungen eingesetzt.

In der vorliegenden Studie sollte die Off-Skala der Überprüfung der konvergenten Validität der SN-S dienen, der Gesamtwert (FAF-Agg) der Überprüfung des Zusammenhangs zwischen Antwortverhalten und testpsychologisch ermittelter Aggressionsbereitschaft. Ausgefüllte FAF-Bögen lagen von 21 Probanden vor (Schuldfähigkeit: $n = 14$, Prognose $n = 6$, Gesundheitsnachweis: $n = 1$). Aus der Akte wurden Roh- und T-Wert des Gesamtwerts und der Off-Skala entnommen.

4.3.6 Rosenzweig Picture-Frustration Test (PFT)

Der PFT (Rosenzweig, 1976; Rauchfleisch, 1979a, 1979b) ist ein semiprojektives Verfahren zur Erfassung der Frustrationstoleranz i.S. der Belastbarkeit einer Persönlichkeit in sozialen Situationen. Er besteht aus einem Testheft mit 24 skizzenartig gezeichneten Situationen, in denen eine von zwei oder mehreren Personen etwas äußert, das bei einer anderen abgebildeten Person potentiell Frustration hervorrufen kann. Die Reaktion dieser Person soll vom Probanden in eine leere Sprechblase eingetragen werden. Dabei soll die Antwort möglichst spontan erfolgen, indem das erste, was einem in den Sinn kommt, eingetragen wird. Die Antworten werden mithilfe eines standardisierten Signierungssystems kodiert. Erfasst werden drei verschiedene Richtungen der Aggressionsäußerung: gegen die Umwelt (Extrapunitivität), gegen die eigene Person (Intropunitivität) und ein Umgehen der Aggression (Impunitivität). Diese Richtungen sind wiederum unterteilt in drei verschiedene Reaktionstypen, nämlich in den Obstacle-Dominance-Typ, wenn der Frustrationscharakter der Situation im Vordergrund steht, den Ego-Defense-Typ, wenn der Bezug auf das Ich des Probanden dominiert, und den Need-Persistence-Typ, wenn explizit eine Lösung der Situation angestrebt wird. Aus der Kombination dieser Kategorien resultieren neun Auswertungsfaktoren, denen zwei weitere hinzugefügt wurden (aggressive Leugnung der eigenen

Verantwortung und Attribution der eigenen Schuld auf unvermeidbare Umstände). Die Anzahl der jeweiligen Signierungen ergibt den Rohwert, der anhand von alters- und geschlechtsspezifischen Tabellen in einen Stanine-Wert umgewandelt werden kann.

Die Auswertungsobjektivität des PFT ist aufgrund des Signierungssystems höher als bei anderen projektiven Verfahren. Im Handbuch (Rauchfleisch, 1979b) werden Beurteilerübereinstimmungen zwischen 72% und 100% berichtet. Die Retest-Reliabilität lag in verschiedenen Studien zwischen .30 und .90 für die Aggressionsrichtungen und Reaktionstypen und zwischen .27 und .83 für die Einzelsignierungen. Im Handbuch werden die Ergebnisse verschiedener Validierungsstudien berichtet, die korrelative Zusammenhänge zwischen den PFT-Signierungen und verschiedenen Außenkriterien untersuchten. Dabei konnte unter anderem ein signifikanter Zusammenhang zwischen den extrapunitiven Antworten und einer niedrigen Aggressionshemmung im Minnesota Multiphasic Personality Inventory sowie hohen Werten auf den Skalen Paranoia und Psychopathie gezeigt werden. Die Befunde zur Vorhersage des tatsächlichen Verhaltens (z.B. manifeste Aggressivität) waren dagegen uneinheitlich.

Obwohl die Validität des Verfahrens umstritten ist (Graybill & Heuvelman, 1993), ist es gerade im forensischen Bereich noch weit verbreitet (vgl. z.B. Lemm, 2000; Littmann, 2007), da es als weniger leicht zu durchschauen und zu verfälschen gilt als entsprechende Fragebögen. Allerdings konnte Vane (1954) zeigen, dass es einen Zusammenhang zwischen Impression Management und den Antworten im PFT gibt und dass delinquente Mädchen häufiger sozial erwünscht antworten als gematchte, nicht-delinquente Kontrollprobandinnen. Auch die widersprüchlichen Befunde der Studien, die die PFT-Antworten von (Gewalt-)Straftätern und Kontrollprobanden verglichen, sprechen für eine Verfälschbarkeit des Verfahrens durch positive Antwortverzerrungen. King (1972) konnte zeigen, dass sozial erwünschtes Antworten bei Straftätern zu einer Zunahme intro-punitiver und einer Abnahme extrapunitiver Antworten führt.

In vorliegender Arbeit wurden jedoch nicht alle intro- und extrapunitiven Antworten erhoben, sondern lediglich die sog. E-Antworten, die eine aggressive Äußerung gegen die Umwelt enthalten (z.B. offener Tadel oder Beschuldigungen). Solche Antworten gelten innerhalb des Normbereichs als Indikator für die Durchsetzungsfähigkeit einer Person, eine überdurchschnittliche Anzahl solcher Reaktionen ist jedoch als Hinweis für eine erhebliche Aggressionsspannung zu sehen bzw. als Maß für die Neigung, „reine“ Aggressionen gegen die Außenwelt zu richten.

Aus der Akte wurden die Stanine-Werte entnommen. Der PFT lag von insgesamt 55 Probanden vor (Schuldfähigkeit: $n = 23$, Prognose $n = 21$, Gesundheitsnachweis: $n = 1$).

4.3.7 Foto-Hand-Test (FHT)

Der FHT (Belschner, Lischke & Selg, 1971) ist ein projektives Verfahren zur Erfassung der Aggressionsbereitschaft. Er besteht aus 34 schwarz-weiß Fotografien, auf denen jeweils eine männliche Hand abgebildet ist. Der Proband wird aufgefordert anzugeben, was die abgebildete Hand seiner Meinung nach gerade tun könnte. Grundsätzliche Annahme der Testkonstruktion war, dass die Handlungstendenzen, die ein Proband den Händen zuordnet, charakteristisch für sein Verhalten sind. Die Signierung der Antworten erfolgt nach einem festgelegten Kategoriensystem, das aus insgesamt 18 Hauptkategorien besteht, von denen jedoch nur die ersten sechs Relevanz für die Bestimmung der Aggressionsbereitschaft besitzen. Dabei handelt es sich einerseits um die aggressiven (agg) Antworten (drohen, angreifen, verletzen) und die direktiven (dir) Antworten (jemanden lenken, beeinflussen), andererseits um die affektiven (aff) Antworten (positive Gefühle gegenüber jemandem äußern), die kommunikativen (komm) Antworten (Informationen austauschen), die dependenten (dep) Antworten (um Hilfe oder Beistand bitten) und die Furcht- (fur) Antworten (Angst, Furcht äußern oder Aggressionen ausgesetzt sein). Durch Subtraktion der kontraaggressiven (komm, aff, dep, fur) von den aggressiven (agg, dir) Antworten wird der sog. Acting-Out-Score (AOS) berechnet, der als Indikator für offen aggressives Verhalten gilt. Für die Interpretation des AOS stehen Normwerte in Form von Mittelwerten und Standardabweichungen von gesunden, unauffälligen sowie delinquenten Jugendlichen und Erwachsenen zur Verfügung, die jedoch aufgrund der geringen Stichprobengröße als ungenügend zu beurteilen sind.

Die Auswertungsobjektivität ist durch das Signierungssystem gegeben (Beurteilerübereinstimmung $r_{12} = .96$, $p < .01$). Die Retest-Reliabilität des AOS ist mit $r_{tt} = .90$ ($KI = .71-.97$) als gut zu beurteilen. Zur Validität liegen eher wenige Daten vor. Die im Manual berichteten Validierungsstudien zeigen u.a., dass heranwachsende und erwachsene Straftäter signifikant höhere AOS-Werte als gematchte Kontrollprobanden erzielen und erwachsene Gewaltstraftäter höhere Werte als nicht-gewalttätige Straftäter. Dagegen zeigte sich kein signifikanter Zusammenhang mit der selbstberichteten Aggressivität im FAF. Die Verfälschbarkeit des Verfahrens wird von den Autoren als gering eingestuft. Der Testzweck konnte nur von wenigen Kontrollprobanden richtig erkannt werden und unterschiedlich instruierte Stichproben mit der Aufforderung, sich entweder in besonders gutem oder besonders

schlechtem Licht darzustellen, unterschieden sich nicht von Probanden, die das Verfahren mit der Standardinstruktion bearbeiteten.

Für die vorliegende Arbeit wurde der AOS-Rohwert und seine Interpretation als über- oder unterdurchschnittlich bzw. normgerecht gemäß den Normwerten des Manuals ($M \pm 1 SD$) aus den Akten entnommen. Er lag von insgesamt 35 Probanden vor (Schuldfähigkeit: $n = 17$, Prognose: $n = 18$).

4.3.8 Skala „Impression Management“ aus dem 16 Persönlichkeitsfaktoren-Test – Revidierte Fassung (16 PF-R)

Der 16 PF-R (Schneewind & Graf, 1998) ist ein Inventar zur mehrdimensionalen Persönlichkeitsdiagnostik in unterschiedlichen Anwendungsfeldern. Er besteht aus 184 Items, die auf einer dreistufigen Skala („stimmt“, „?“ „stimmt nicht“) einzuschätzen sind. Es können 16 Primärdimensionen und fünf Globalfaktoren (Extraversion, Unabhängigkeit, Selbstkontrolle, Ängstlichkeit, Unnachgiebigkeit) sowie eine Skala zur Erfassung sozial erwünschten Antwortverhaltens (Impression Management – IM; s. Anhang A5) abgeleitet werden. Außerdem können zwei weitere Antwortstil-Indizes berechnet werden (Akquieszenz und Infrequenz). Die IM-Skala umfasst 10 Items, der Summenwert kann anhand von alters- und geschlechtsspezifischen Tabellen in Stenwerte transformiert werden. Überdurchschnittliche Werte (Stenwert > 6) sprechen für eine übermäßige Tendenz, sich selbst in möglichst günstigem Licht zu präsentieren.

Gütekriterien werden für die Kontrollskalen im Handbuch nicht berichtet, aber Korrelationen der IM-Skala mit den einzelnen Primärdimensionen. Probanden mit sozial erwünschtem Antwortverhalten beschreiben sich demnach vor allem als emotional stabil, selbstsicher, entspannt, regelbewusst, bodenständig und perfektionistisch.

Der 16 PF-R wird bei Begutachtungen eingesetzt, um das allgemeine Persönlichkeitsbild von Probanden zu erfassen. Im Gegensatz zu vergleichbaren Verfahren wie dem NEO-Persönlichkeitsinventar nach Costa & McCrae (Ostendorf & Angleitner, 2004) ist sein Einsatz aufgrund der komplexen Auswertestrategie weniger ökonomisch, wegen der Integration von Antwortkontrollskalen wird ihm dennoch bisweilen der Vorzug gegeben.

Die IM-Skala diente in der vorliegenden Arbeit der Überprüfung der konvergenten Validität. Ausgefüllte Fragebögen lagen von 10 Probanden vor (Schuldfähigkeit: $n = 5$, Prognose $n = 5$). Aus der Akte wurden jeweils Roh- und Stenwert entnommen.

4.3.9 Skala „Offenheit“ aus dem Eppendorfer Schizophrenie-Inventar (ESI)

Das ESI (Maß, 2001) ist ein Selbstbeurteilungsverfahren, das der quantitativen Erfassung subjektiver schizophrenietypischer Phänomene dient, wie z.B. Aufmerksamkeits- und Wahrnehmungsstörungen sowie psychotische Symptome. Ziel der Konstruktion war es, ein Instrument zur Abbildung des unmittelbaren Erlebens des psychotischen Prozesses zu entwickeln. Die Skala Offenheit (ESI-FR) misst die Bereitschaft, kleinere Verstöße gegen sozial erwünschte Verhaltensregeln einzugestehen (s. Anhang A6). Sie soll laut Autoren außerdem Aufschluss über das Vorhandensein von Verleugnungstendenzen gegenüber der Krankheit geben. Sie besteht aus 5 Items, die auf einer vierstufigen Skala von 0-3 Punkten (von „stimmt genau“ bis „stimmt gar nicht“) einzuschätzen sind. Der Skalenrohwert entspricht der Summe aller Items. Im Manual werden Normwerte (z-, T- und Stanine-Werte) von einer Stichprobe schizophrener Patienten, einer gemischten klinischen Vergleichsgruppe akut-stationärer psychiatrischer Patienten sowie einer Gruppe gesunder Kontrollprobanden zur Verfügung gestellt. Werden auf der Kontrollskala weniger als zwei Punkte erreicht, gelten die Ergebnisse der Inhaltsskalen als fragwürdig. Die interne Konsistenz der FR-Skala wird mit Cronbachs $\alpha = .73$, die Retest-Reliabilität mit $r_{tt} = .85$ bzw. $r_{tt} = .86$ angegeben.

Das ESI kam ausschließlich bei Begutachtungen von Probanden zum Einsatz, bei denen in der Vorgeschichte eine Schizophrenie diagnostiziert worden war. Für die Interpretation der Skalen wurden die Normwerte der gesunden Kontrollprobanden herangezogen, wobei ab einem T-Wert < 40 das Vorliegen positiver Antwortverzerrungen angenommen wurde. Die ESI-FR lag von 9 Probanden vor (Schuldfähigkeit: $n = 4$, Prognose $n = 5$).

4.3.10 Kontrollskalen des Minnesota Multiphasic Personality Inventory – 2 (MMPI-2)

Das MMPI-2 (Engel, 2003) ist einer der weltweit am häufigsten eingesetzten klinischen Persönlichkeitsfragebögen. Es besteht aus 567 Items, die auf einer zweistufigen Antwortskala als richtig oder falsch einzuschätzen sind. Das Verfahren dient der Erfassung der Persönlichkeitsstruktur insbesondere von Personen mit psychischen Problemen und Störungen. Neben den verschiedenen Inhaltsskalen verfügt das Instrument über drei Kontrollskalen, die Auskunft über die Gültigkeit des gewonnenen Profils geben (s. Anhang A7). Die sog. Lügen-(L)-Skala erfasst die Tendenz, die Items so zu beantworten, dass systematisch nicht die eigene, sondern eine „ideale“ oder „perfekte“ Person beschrieben wird. Erhöhte Werte (T-Wert: 60-69) sprechen für eine abwehrende Haltung gegenüber der Untersuchung, hohe Werte (T-Wert: 70-79) für ein Leugnen eigener Fehler oder eine zufällige

Beantwortung und sehr hohe Werte (T-Wert > 79) für eine Dissimulation. Sehr niedrige Werte (T-Wert < 50) können dagegen auf eine Übertreibung von Problemen und Auffälligkeiten hinweisen. Die Seltenheit-(F)-Skala dient der Identifikation von Probanden, die die Items nach dem Zufallsprinzip beantworten, entweder aufgrund mangelnder Kooperation oder wegen mangelnder Intelligenz, einer Lesestörung, mangelndem Realitätskontakt oder Aufmerksamkeitsstörungen. Hohe Werte (T-Wert: 71-90) können sowohl durch eine psychotische Störung als auch durch eine starke Zustimmungstendenz oder eine Simulation bedingt sein. Sehr hohe Werte (T-Wert > 90) sprechen für eine Beantwortung nach dem Zufall oder eine schwere Leseschwäche. Sehr niedrige Werte (T-Wert < 45) können Hinweis auf ein konformistisches Antwortverhalten und eine mögliche Dissimulation sein. Die komplexeste Validitätsskala ist die Korrektur-(K)-Skala. Sehr hohe Werte (T-Wert > 70) sprechen für eine abwehrende Haltung, Leugnung von Problemen und Dissimulationsneigung, können aber auch von gesunden Personen ohne psychische Probleme erzielt werden. Niedrige Werte (T-Wert < 41) sind dagegen ein Zeichen für Simulation bzw. eine extreme Zustimmungstendenz.

Das MMPI-2 wurde an einer großen, bevölkerungsrepräsentativen Stichprobe normiert, es liegen geschlechtsspezifische Normwerte in Form von T-Werten vor. Die Reliabilität der drei Kontrollskalen ist ausreichend bis gut. Die Retest-Reliabilität liegt zwischen $r_{tt} = .71$ und $r_{tt} = .90$, die interne Konsistenz zwischen $\alpha = .60$ und $\alpha = .81$. Die Validität der Skalen der englischen Originalversion wurde in einer Reihe von Studien (Übersicht bei Greene, 2008), auch an forensischen Populationen (z.B. Bagby, Rogers & Buis, 1994; Sellbom, Toomey, Wygant, Kucharski & Duncan, 2010) überprüft.

Das MMPI-2 wird aufgrund seines Umfangs nur selten bei Begutachtungen eingesetzt, und nur dann, wenn das Vorliegen einer schweren psychischen Störung, wie Schizophrenie oder Depression, testpsychologisch untersucht werden soll. Die Auswertung erfolgt mithilfe des faxgestützten Auswertungs- und Interpretationssystems des Hogrefe Verlags. Dieses liefert neben den numerischen Skalenwerten auch eine computergenerierte Profilinterpretation. Für die vorliegende Untersuchung wurden jedoch lediglich Roh- und T-Werte der drei Kontrollskalen aus der Akte entnommen. Die MMPI-Kontrollskalen lagen von 11 Probanden vor (Schuldfähigkeit: $n = 4$, Prognose $n = 2$, Zahlungsbegehren: $n = 3$, Gesundheitsnachweis: $n = 2$).

4.3.11 Strukturierter Fragebogen Simulierter Symptome (SFSS)

Der SFSS (Cima et al., 2003b) ist die deutsche Übersetzung des Structured Inventory of Malingered Symptomatology (Smith & Burger, 1997). Es handelt sich um einen Screening-Fragebogen zur Erfassung von Aggravation bzw. Simulation von psychiatrisch relevanten Symptomen. Das Instrument besteht aus 75 Items (s. Anhang A8), die auf einer zweistufigen Antwortskala im Ja-Nein-Format als wahr oder falsch zu beurteilen sind. Je 15 Items sind einer von fünf Skalen zugeordnet: Affektive Störungen (AS), Amnestische Störungen (AM), Neurologische Beeinträchtigung (N), Niedrige Intelligenz (NI) und Psychose (P). Der jeweilige Skalenrohwert entspricht der Anzahl auffällig beantworteter Items, der SFSS-Gesamtwert der Gesamtzahl solcher Items. Für alle Skalen und den Gesamtwert liegen Cut-off-Werte vor. Eine Überschreitung dieser Werte spricht für das Vorliegen negativer Antwortverzerrungen. Die Studie von Cima et al. (2003b), in welcher forensische Psychiatriepatienten sowie unterschiedlich instruierte und ehrlich antwortenden Studenten verglichen wurden, erbrachte zufriedenstellende Kennwerte für die deutsche Version. Die Retest-Reliabilität lag bei $r_{tt} = .97$, die interne Konsistenz bei Cronbachs $\alpha = .80$. Die Sensitivität des Gesamtwerts und der negative Vorhersagewert betrugen jeweils 87%, Spezifität und positiver Vorhersagewert jeweils 86%. Unter den forensischen Probanden hatten 29% einen auffälligen Gesamtwert erzielt, von ihnen waren 65% von unabhängigen Beurteilern nach Analyse der Krankengeschichte als Simulanten eingestuft worden. Dass Patienten mit psychiatrischen Störungen im SFSS bisweilen fälschlicherweise auffällige Werte erzielen, konnte auch in zwei eigenen Studien gezeigt werden, in denen kleine Gruppen stationär-psychiatrischer Patienten der LMU ($N = 17$ bzw. $N = 30$) mit dem SFSS untersucht wurden (Mader, 2014; Veress, 2010). Hier erzielten 29% bzw. 50% der Patienten auffällige Werte. Allein aufgrund eines auffälligen Wertes darf daher nicht auf einen bewussten Täuschungsversuch geschlossen werden (Dressing, Foerster, Widder, Schneider & Falkai, 2011), dieser könnte auch durch unbewusste Aggravationstendenzen im Rahmen bestimmter psychischer Störungen oder bei einer Begutachtung erwartbare Verdeutlichungstendenzen bedingt sein kann (Rogers, 2008a; Widder, 2007).

Der SFSS wird bei Begutachtungen immer dann eingesetzt, wenn das Vorliegen negativer Antwortverzerrungen ausgeschlossen werden soll, also bei allen Probanden mit Zahlungsbegehren und i.d.R. bei Schuldfähigkeitsbegutachtungen. Ausgefüllte Bögen lagen daher von allen Probanden dieser beiden Gruppen vor. Bei den anderen Begutachtungen wurde er seltener eingesetzt und lag von 37 Prognoseprobanden und 6 Gesundheitsnachweis-

Probanden vor. Aus der Akte wurden sowohl die Einzelitems als auch Skalen- und Gesamtwerte entnommen.

4.4 Statistische Analyse der Daten

Die statistische Auswertung der Daten erfolgte mit den Programmen IBM® SPSS für Windows (Version 20) und Microsoft® Office Excel 2007.

Für alle Signifikanztests wurde ein Signifikanzniveau von $p < .05$, zweiseitig, festgelegt.

Auch bei Verletzung der jeweiligen Voraussetzungen (z.B. Normalverteilung) wurden bei Mittelwertsvergleichen parametrische Verfahren (Varianzanalyse, t-Tests) eingesetzt, da sie gegenüber solchen Verletzungen recht robust sind und eine höhere Power besitzen als äquivalente non-parametrische Verfahren (Bühner, 2011). Zusammenhänge zwischen kategorialen Variablen wurden mit Kreuztabellen und χ^2 - bzw. Fishers exakten Tests, wenn mehr als 20% der Zellen eine erwartete Häufigkeit kleiner 5 aufwiesen, untersucht.

Für die Auswertungen wurden entweder die Gesamtstichprobe oder bestimmte Substichproben herangezogen, außerdem unterschiedliche Rekodierungen der einzelnen Variablen vorgenommen. Das Vorgehen im Einzelnen wird im Folgenden beschrieben.

– Reliabilitätsanalyse

Zur Schätzung der Reliabilität wurde die interne Konsistenz (Cronbachs α) des SN-S-Gesamtwerts und der Subskalen jeweils für die Gesamtstichprobe und alle Substichproben berechnet. Eine Skala Angst konnte nicht gebildet werden, da diese nur aus einem Item besteht. Mit der Reliabilitätsanalyse wurden auch jeweils eine Itemanalyse durchgeführt und der jeweilige Itemmittelwert (als Schwierigkeitsindex), die Standardabweichung, die mittlere Inter-Item-Korrelation (MIC) als Homogenitätsindex der jeweiligen Skala und die Itemtrennschärpen berechnet. Es wurde ein listenweiser Fallausschluss vorgenommen, sodass Personen mit mindestens einem fehlenden Wert aus der folgenden Reliabilitätsanalyse ausgeschlossen wurden. Dieselbe Prozedur wurde auch für die PPI-Skalen UA und SE und später für die durch die Faktorenanalyse extrahierten Skalen (s.u.) durchgeführt.

– Faktorenanalyse

Um die faktorielle Validität zu überprüfen, wurden die Daten der Gesamtstichprobe herangezogen. Es wurde eine Hauptkomponentenanalyse mit obliquen Rotationsmethode (Promax) durchgeführt, die Dummy-Items wurden dabei nicht berücksichtigt und es

wurde ein listenweiser Fallausschluss vorgenommen. Da die SN-S aus sieben Subskalen besteht, wurden sieben Faktoren extrahiert.

Die extrahierten Faktoren wurden weiteren Analysen unterzogen (s.u.). Durch den listenweisen Fallausschluss wurden Personen, bei denen mindestens ein Item auf diesem Faktor fehlte, nicht in die weiteren Analysen eingeschlossen.

– *Bestimmung der Cut-off-Werte und der diagnostischen Validität der SN-S*

Für diese Analysen wurden die Daten von KG und Fake-KG herangezogen. Der optimale Cut-off-Punkt für den SN-S-Gesamtwert und die Subskalen wurde mit Receiver Operating Characteristic- (ROC-) Analysen bestimmt, wobei die Gruppenzugehörigkeit zur Fake-KG als Kriterium (Supernormalität vorhanden) diente. Als idealer Cut-off wurde der Punktwert gewählt, bei dem der Youden-Index (Sensitivität + Spezifität – 1; Goldhammer & Hartig, 2012) den höchsten Wert erreichte.

Die Bestimmung der Validität erfolgte zum einen über die Berechnung der Area Under Curve (AUC). Für die Berechnung der in Kapitel 2.4.1 beschriebenen Indizes wurde zunächst das Antwortverhalten der Probanden entsprechend des für die jeweilige Skala errechneten Cut-off-Werts in unauffällig und supernormal klassifiziert. Dann wurden die Werte entsprechend der in Anhang B aufgeführten Formeln berechnet.

Dasselbe Procedere wurde für die aus der Faktorenanalyse extrahierten Skalen und die PPI-SE-Skala angewendet. Außerdem wurde eine erste Kreuzvalidierung vorgenommen, indem die Güteindizes der diagnostischen Validität auch anhand der KV-Stichprobe bestimmt wurden.

– *Bestimmung der konvergenten Validität*

Um die dimensionale Übereinstimmung mit anderen Kontrollskalen zu prüfen, wurden Produkt-Moment-Korrelationen bzw. Spearman-Rangkorrelationen (bei $N < 30$) von SN-S-Gesamtwert und Subskalen mit anderen Skalen zur Erfassung positiver Antwortverzerrungen (PPI-UA, PPI-SE, FAF-Off, PFR-IM, ESI-FR) berechnet. Um die kategoriale Übereinstimmung zu prüfen, wurden die Skalenwerte anhand der jeweiligen Cut-off- bzw. Normwerte in dichotome Variablen mit den Stufen auffälliges vs. unauffälliges Antwortverhalten rekodiert. Für die drei MMPI-Kontrollskalen, bei denen eine Dichotomisierung nicht möglich ist, wurden mehrstufige kategoriale Variablen gebildet. Die konvergente Validität wurde dann mit Kreuztabellen bestimmt.

– *Erfassung von negativen Antwortverzerrungen mit der SN-S*

Zunächst wurden Korrelationen zwischen SN-S und SFSS (jeweils Gesamtwert und Subskalen) berechnet. Um zu überprüfen, ob hohe Werte in der SN-S für das Vorliegen

negativer Antwortverzerrungen sprechen, wurden ROC-Analysen für den Gesamtwert berechnet. Als Außenkriterium für das Vorliegen negativer Antwortverzerrungen wurde das Ergebnis im SFSS (> 16 Punkte) herangezogen. Mithilfe des Youden-Index wurde der optimale Cut-off-Wert bestimmt und anschließend die Validität der Klassifikation nach dem oben beschriebenen Procedere bestimmt.

– *Häufigkeit von Antwortverzerrungen bei forensischen Begutachtungsprobanden*

Um zu überprüfen, ob supernormales Antwortverhalten in Abhängigkeit von der Fragestellung bzw. dem Zweck der Begutachtung unterschiedlich häufig auftritt, wurden die verschiedenen Substichproben miteinander verglichen. Dabei wurde zum einen wie von Cima et al. (2008) ein dimensionaler Vergleich vorgenommen, indem mit dem Faktor Gruppenzugehörigkeit einfaktorielle Varianzanalysen für die SN-S sowie Bonferroni Post hoc Tests berechnet wurden. Zum anderen wurden mit der kategorialen Einstufung des Antwortverhaltens als auffällig oder unauffällig Kreuztabellen berechnet. Für die Kreuztabellen wurden post hoc jeweils zwei Gruppen miteinander verglichen. Wegen der Vielzahl an Einzelvergleichen wurde eine Bonferroni-Korrektur durchgeführt. Das korrigierte Signifikanzniveau war $p < .002$. In die Analyse wurden jeweils auch KG und Fake-KG einbezogen, um zu untersuchen, ob sich die Begutachtungsgruppen von diesen unterscheiden.

– *Antwortverhalten und testpsychologisch ermittelte Aggressionsbereitschaft*

Zunächst wurden Produkt-Moment- bzw. Spearman-Rangkorrelationen (bei $N < 30$) zwischen den drei Verfahren zur Erfassung der Aggressionsbereitschaft und dem SN-S-Gesamtwert sowie den Skalen zur Minimierung von Aggression berechnet. Dann wurden der dimensionale (t-Tests) und der kategoriale (χ^2 - bzw. Fishers exakten Tests) Zusammenhang zwischen SN-S-Gesamtwert sowie den Subskalen, die eine Minimierung von Aggressivität erfassen, und der mit den testpsychologischen Verfahren gemessenen Aggressionsbereitschaft untersucht. Für die kategoriale Übereinstimmung wurden FAF-Agg, FHT-AOS und die Anzahl der E-Antworten im PFT anhand der jeweiligen Normwerte in kategoriale Variablen rekodiert (unterdurchschnittliche, durchschnittliche und überdurchschnittliche Aggressionsbereitschaft).

Dann wurden kategoriale Variablen gebildet, durch welche die Übereinstimmung bzw. Diskrepanz zwischen selbstberichteter Aggressivität im FAF und der Aggressionsbereitschaft in PFT und FHT kodiert wurde. Eine vorhandene Übereinstimmung wurde kodiert, wenn in beiden Verfahren ein durchschnittlicher, ein unterdurchschnittlicher oder ein überdurchschnittlicher Wert erzielt wurde. Wenn im projektiven Verfahren ein

überdurchschnittlicher, im FAF aber ein (unter-)durchschnittlicher Wert erreicht wurde, wurde das Label „Dissimulation“ vergeben. Außerdem wurde eine „Aggravation“ kodiert, wenn im projektiven Verfahren ein (unter-)durchschnittlicher und im FAF ein überdurchschnittlicher Wert erreicht wurde. Anhand von Kreuztabellen wurde die Übereinstimmung dieser kategorialen Variablen mit dem dichotomen Werten des SN-S-Gesamtwerts und den Subskalen zur Minimierung von Aggression berechnet.

- *Antwortverhalten und unterdurchschnittliche Skalenwerte in klinischen Persönlichkeitsinventaren*

Der Zusammenhang zwischen supernormalem Antwortverhalten (SN-S-Gesamtwert) und der Anzahl unterdurchschnittlicher Skalen in den Inventaren PSSI, IKP-G und -Eg wurde mithilfe von Spearman-Rangkorrelationen und t-Tests untersucht.

- *Antwortverhalten und Persönlichkeitsstörung*

Zunächst wurden anhand von t-Test und Kreuztabellen die SN-S-Werte von Probanden mit und ohne Persönlichkeitsstörung verglichen. Dann wurde die Übereinstimmung zwischen selbstberichteter und diagnostizierter Persönlichkeitsstörung anhand einer kategorialen Variable kodiert. Dabei wurden nur die Daten von Personen mit diagnostizierter Persönlichkeitsstörung berücksichtigt. Eine Übereinstimmung wurde kodiert, wenn der Proband auf den der Diagnose entsprechenden Skalen der Persönlichkeitsinventare PSSI oder IKP-G einen überdurchschnittlichen Wert erzielte, eine Dissimulation, wenn er auf den entsprechenden Skalen (unter-)durchschnittliche Werte erzielte. Der Zusammenhang mit supernormalem Antwortverhalten wurde dann mithilfe von Kreuztabellen untersucht.

5 Ergebnisse

5.1 Item-Statistiken und Reliabilitätsanalyse

Die interne Konsistenz des SN-S-Gesamtwerts (s. Tabelle 6) fiel in der Gesamtstichprobe ($N = 281$) mit $\alpha = .92$ sehr gut aus. Die Ergebnisse der Einzelitemanalysen können dem Anhang entnommen werden (Anhang C, Tabelle 25). Auch in den Stichproben Schuldfähigkeit ($\alpha = .95$), Prognose ($\alpha = .92$) und Zahlungsbegehren ($\alpha = .95$) wurden jeweils sehr gute Werte erreicht. In der Gruppe Gesundheitsnachweis konnte die interne Konsistenz des Gesamtwerts und auch weiterer Skalen nicht berechnet werden, weil sechs Items keine Varianz aufwiesen, der Mittelwert betrug jeweils 1.0. In der Kontrollgruppe ($\alpha = .86$) hatte der SN-S-Gesamtwert gute Reliabilitätswerte, in der Fake-KG ($\alpha = .73$) zeigte er einen noch akzeptablen Wert. Hier lag die mit Abstand geringste mittlere Inter-Item-Korrelation vor, die mit $MIC = .08$ unzureichend ausfiel. Die Items wiesen in dieser Gruppe die niedrigsten Mittelwerte und geringsten Standardabweichungen auf.

Tabelle 6

Interne Konsistenz (Cronbachs α) und mittlere Inter-Item-Korrelation (MIC) des SN-S-Gesamtwerts und der Subskalen in der Gesamtstichprobe (GS) und den einzelnen Substichproben (Fortsetzung auf S. 59)

SN-S	GS	SF	RP	ZB	GN ^a	KG	Fake-KG
Gesamtwert							
<i>N</i>	281	33	42	20	-	114	65
Cronbachs α	.92	.95	.92	.95	-	.86	.73
MIC	.26	.35	.24	.34	-	.15	.08
Soziale Erwünschtheit							
<i>N</i>	304	34	42	22	8	128	70
Cronbachs α	.52	.62	.47	.41	.32	.53	.10
MIC	.21	.29	.19	.12	.09	.22	.06

Fortsetzung Tabelle 6	GS	SF	RP	ZB	GN ^a	KG	Fake-KG
Affektive Symptome							
<i>N</i>	301	33	43	22	-	127	69
Cronbachs α	.84	.87	.87	.89	-	.75	.64
MIC	.43	.48	.50	.55	-	.31	.21
Zwang							
<i>N</i>	305	34	43	23	-	128	69
Cronbachs α	.61	.71	.54	.67	-	.46	.20
MIC	.29	.39	.27	.33	-	.17	.06
Dissoziatives Erleben							
<i>N</i>	303	34	43	22	-	127	69
Cronbachs α	.61	.73	.32	.83	-	.45	.45
MIC	.25	.37	.12	.51	-	.15	.16
Psychotische Symptome							
<i>N</i>	306	34	43	23	-	128	70
Cronbachs α	.73	.90	.75	.81	-	.58	.37
MIC	.33	.62	.35	.45	-	.20	.11
Aggression							
<i>N</i>	307	34	43	23	-	130	69
Cronbachs α	.78	.83	.81	.87	-	.67	.57
MIC	.35	.41	.37	.51	-	.23	.19

Anmerkung: ^aIn dieser Stichprobe konnte wegen fehlender Varianz bei sechs Items nur für die Skala Soziale Erwünschtheit die interne Konsistenz berechnet werden.

Die einzelnen Skalen der SN-S wiesen deutlich schlechtere Reliabilitäten auf (s. Tabelle 6). Lediglich die Skalen Affektive Symptome, Psychotische Symptome und Aggression erreichten in der Gesamtstichprobe und in den meisten Substichproben akzeptable bis gute Werte ($\alpha > .70$). Die Skala Dissoziatives Erleben hatte nur in den Stichproben Zahlungsbe-

gehen und Schuldfähigkeit eine ausreichende interne Konsistenz, die Skala Zwang in der Schuldfähigkeitsgruppe. Die Skala soziale Erwünschtheit wies dagegen in allen Stichproben fragwürdige bis unzureichende Werte auf ($\alpha < .70$).

Die interne Konsistenz der PPI-UA-Skala konnte für die Stichprobe der Prognoseprobanden und die Fake-KG nicht berechnet werden, weil ein bzw. zwei Items keine Varianz aufwiesen, der Mittelwert betrug jeweils 1.0. Die Skala hatte sowohl in der Gesamtstichprobe als auch in den Teilstichproben der Schuldfähigkeitsprobanden und der KG schlechte bis unzureichende Reliabilitätswerte ($\alpha = .31-.67$, s. Tabelle 7). Die PPI-SE-Skala wies in der Gesamtstichprobe und den Teilstichproben mit Ausnahme der Stichprobe der Prognoseprobanden eine akzeptable interne Konsistenz auf ($\alpha = .71-.75$). Die mittleren Inter-Item-Korrelationen waren höher als die der Skala PPI-UA und lagen in der Gesamtstichprobe und der Fake-KG im akzeptablen Bereich (s. Tabelle 7). Die weiteren Ergebnisse der Itemanalyse können Tabelle 26 in Anhang C entnommen werden.

Tabelle 7

Interne Konsistenz (Cronbachs α) und mittlere Inter-Item-Korrelation (MIC) der PPI-R Skalen Unaufrichtige Beantwortung und Soziale Erwünschtheit in der Gesamtstichprobe (GS) und den einzelnen Substichproben

PPI-R	GS	SF	RP ^a	KG	Fake-KG ^a
Unaufrichtige Beantwortung					
<i>N</i>	221	19	-	117	-
Cronbachs α	.67	.31	-	.58	-
MIC	.08	.02	-	.07	-
Soziale Erwünschtheit					
<i>N</i>	229	19	22	122	66
Cronbachs α	.77	.70	.68	.71	.75
MIC	.24	.17	.18	.18	.22

Anmerkung: ^aIn diesen Stichproben konnte wegen fehlender Varianz bei einem (RP) bzw. zwei (Fake-KG) Items die interne Konsistenz für die Skala PPI-UA nicht berechnet werden.

Einige Items der SN-S und der PPI-R-Skalen wiesen unzureichende Trennschärfen ($r < .30$), einige Items der PPI-R-Skalen negative Trennschärfen auf. Da eine Revision der Ska-

len aber nicht das Ziel der Analysen war, wird auf diese Ergebnisse hier nicht näher eingegangen.

5.2 Faktorenanalyse

In die Analyse wurden $N = 281$ Probanden eingeschlossen. Nach Ausschluss der Dummy-Items zeigte sich eine gute Eignung der SN-S-Items für eine Faktorenanalyse. Es lagen substantielle Korrelationen vor ($KMO = .91$), alle MSA-Koeffizienten (Partialkorrelationen des Items mit den übrigen Items) waren $> .50$ und der Bartlett-Test auf Sphärizität, der prüft, ob die Korrelationen der Korrelationsmatrix signifikant von Null verschieden sind, war signifikant ($p < .001$). Nach Extraktion von sieben Faktoren wiesen alle Items Kommunalitäten von $h^2 > .10$ auf. Alle sieben Faktoren hatten einen Eigenwert > 1 . Sie erklärten zusammen 57% der Varianz. Die Analyse der Mustermatrix (Tabelle 8) zeigte, dass lediglich ein Item (19) keine bedeutsame Ladung ($> .30$) auf einem der Faktoren aufwies. Item 9 wies die höchste Ladung auf Faktor 4 auf, allerdings war diese negativ. Eine weitere bedeutsame Ladung wies es auf Faktor 7 auf. Die extrahierten Faktoren ließen sich nur zum Teil inhaltlich sinnvoll interpretieren. Mit Faktor 1 werden am ehesten affektive bzw. depressive Beschwerden erfasst, mit Faktor 2 Aggressivität, mit Faktor 3 psychotisches bzw. dissoziatives Erleben. Faktor 4 ließ sich am ehesten mit Angst und Zwang beschreiben. Auf Faktor 5 und 6 luden jeweils nur zwei Items, eine sinnvolle Skalenbildung war daher nicht möglich. Auch auf Faktor 7 luden lediglich drei Items, mit denen am ehesten soziale Erwünschtheit erfasst wird.

Tabelle 8

Faktorladungen der SN-S Items bei einer Hauptkomponentenanalyse mit Promax-Rotation

(Fortsetzung auf S. 63-64)

SN-S-Item	Komponente						
	1	2	3	4	5	6	7
Ich leide unter Stimmungsschwankungen (15)	.86		-.13				
Hin und wieder bin ich traurig, ohne eigentlich zu wissen warum (33)	.82	-.12		.17			
Manchmal fühle ich mich weder körperlich noch psychisch gut (39)	.74	.19			-.11		
Manchmal werde ich von düsteren Gefühlen und Gedanken überwältigt (12)	.68						
Mitunter habe ich mich schon so schlecht gefühlt, dass ich meinte, es könnte nicht so weitergehen (36)	.67	.20	-.16		.12		
Manchmal wird mein Herzschlag schneller und ich weiß nicht warum (21)	.61		.35			-.18	-.29
Manchmal fühle ich mich durcheinander (5)	.61					.18	
Es kommt vor, dass ich Schwierigkeiten habe, mich zu konzentrieren (23)	.55		.28		-.15		
Manchmal verbiete ich mir, an etwas bestimmtes zu denken (31)	.53	-.20				-.15	.26
Ich ärgere mich manchmal so sehr über jemanden, dass ich ihm den Hals umdrehen könnte (26)		.74					-.11

Fortsetzung Tabelle 8	Komponente						
	1	2	3	4	5	6	7
Es wird auch in der Zukunft Situationen geben, in denen ich aggressiv reagiere (49)		.72					
Wenn ich wütend bin, kann ich entsetzlich fluchen (42)		.66			-.17		.21
Manchmal möchte ich jemanden schlagen (16)		.66				.12	-.15
In einigen Situationen verliere ich die Beherrschung (7)	.17	.63		-.16	.14		
Ab und zu werfe ich mit Gegenständen, um mich abzureagieren (38)	-.17	.55		.22	.44	-.16	.14
Wenn ich schlecht gelaunt bin, werde ich gereizt und reagiere manchmal ärgerlich (2)	.39	.42	-.12	.22	-.25	.11	
Es kommt vor, dass ich sehr seltsame Gedanken und Fantasien habe (48)			.80		.18		
Meine Fantasien erscheinen mir manchmal real (18)	-.11	-.14	.74		.16	.25	-.16
Ich habe seltsame Fantasien (22)	.23	-.11	.66	-.12	.16	.11	-.10
Wenn ich alleine bin, spreche ich manchmal mit mir selbst (46)		.27	.65	.11	-.32	-.14	
Manchmal bin ich von etwas so gefesselt, dass ich nicht mehr weiß, wo ich bin (25)			.51			.20	
Es kommt vor, dass ich so enthusiastisch bin, dass mir die Gedanken sehr schnell durch den Kopf gehen (44)			.34		-.12	.21	.28
Es kommt vor, dass ich beim Blick in den Spiegel anders als normalerweise aussehe (19)			.28	.26	.25	.15	

Fortsetzung Tabelle 8	Komponente						
	1	2	3	4	5	6	7
Wenn ich nachts auf der Straße gehe, habe ich das Gefühl, verfolgt zu werden (1)	.11	-.12		.72	.16		.19
Wenn ich nachts in einem großen Haus alleine bin, glaube ich manchmal, Schritte zu hören (40)		.21		.68	.14		
Es kommt vor, dass ich überprüfen muss, ob ich die Tür geschlossen habe (13)			.45	.51		-.18	
Es kommt vor, dass ich Stimmen höre, die in Wirklichkeit nicht da sind (29)				.18	.75		
Ich neige zu Zwangshandlungen (35)			.20		.46	-.11	.16
Ich stelle mir manchmal vor, woanders zu sein (11)		.15	.18		-.14	.75	
Manchmal stelle ich mir vor, jemand anders zu sein (8)	.14	-.14			.19	.68	
Ohne die Hilfe anderer wäre das Leben beschwerlicher (4)			-.17				.81
Es gibt wichtige Ereignisse in meinem Leben, an die ich mich nicht erinnern kann (32)	.33	-.15		-.15	.23	-.15	.53
Ich habe schon einmal etwas Verbotenes getan (9)		.29	.15	-.61			.41

Sowohl aus statistischen als auch aus inhaltlichen Gründen war deshalb lediglich eine Bildung von drei Subskalen sinnvoll (Komponente 1-3). Diese können als Minimierung depressiver Beschwerden (MDe), Minimierung von Aggression (MA) und Minimierung von dissoziativem und psychotischem, also ungewöhnlichem Erleben (MUE) bezeichnet werden. Die MDe-Skala enthält alle Items der Skala Affektive Symptome (5, 12, 15, 23, 33, 36) mit Ausnahme des Items 44, außerdem das Item der Skala Angst (21), ein Item der Skala Zwang (31) und ein Item der Skala Soziale Erwünschtheit (39). Die MA-Skala umfasst alle Items der Skala Aggression (2, 7, 16, 26, 28, 38, 42) sowie ein Item der Skala

Soziale Erwünschtheit (49). Die MUE-Skala setzt sich aus zwei Items der Skala Psychotische Symptome (18, 22), zwei Items der Skala Dissoziatives Erleben (25, 46), einem Item der Skala Affektive Symptome (44) sowie einem Item der Skala Zwang (48) zusammen. Diese drei Skalen wiesen lediglich in der Fake-KG, die dritte Skala auch in der KG, unzureichende Reliabilitätswerte auf. In der Gesamtstichprobe und den anderen Substichproben wurden jeweils zufriedenstellende bis sehr gute Werte erreicht (s. Tabelle 9).

Tabelle 9

Interne Konsistenz (Cronbachs α) der aus der Faktorenanalyse extrahierten Skalen in der Gesamtstichprobe (GS) und den einzelnen Substichproben

SN-S-Skala	GS	SF	RP	ZB	KG	Fake-KG
Minimierung depressiver Beschwerden						
<i>N</i>	301	33	43	23	126	68
Cronbachs α	.88	.89	.87	.92	.83	.67
Minimierung von Aggression						
<i>N</i>	307	34	43	23	130	69
Cronbachs α	.82	.85	.85	.89	.73	.59
Minimierung von ungewöhnlichem Erleben						
<i>N</i>	303	34	43	22	128	69
Cronbachs α	.76	.79	.71	.87	.70	.56

Anmerkung: In der Stichprobe Gesundheitsnachweis konnte wegen fehlender Varianz bei einigen Items die interne Konsistenz nicht berechnet werden.

5.3 Berechnung der Cut-off-Werte und diagnostische Validität

5.3.1 SN-S-Gesamtwert

Der ideale Cut-off-Wert für die Kategorisierung des Antwortverhaltens lag bei 61 Punkten. Der Gesamtwert ermöglichte eine valide Einstufung, die AUC war mit .86 ($p < .001$, 95% KI [.81, .91]) als sehr gut zu bezeichnen.

Tabelle 10

Übereinstimmung der Klassifikation des Antwortverhaltens mit dem SN-S-Gesamtwert und der Untersuchungsgruppe

		SN-S-Gesamtwert dichotom		Σ
		unauffällig	supernormal	
Untersuchungs- gruppe	KG (%)	90 (68.7)	41 (31.3)	131 (100)
	Fake-KG (%)	4 (5.7)	66 (94.3)	70 (100)
	Σ (%)	94 (46.8)	107 (53.2)	201 (100)

Insbesondere die Sensitivität (94%) und der negative Vorhersagewert (96%) waren sehr hoch, die Spezifität (69%) und der positive Vorhersagewert (62%) dagegen unzureichend. Die Gesamttrefferquote war mit 78% aufgrund des hohen Anteils falsch Positiver nicht zufriedenstellend (s. Tabelle 10).

5.3.2 SN-S-Subskalen

Für die Skala Soziale Erwünschtheit lag der ideale Cut-off-Wert bei 11 Punkten. Die AUC war mit .77 ($p < .001$, 95% KI [.70, .83]) zwar als zufriedenstellend zu beurteilen, die Güteindizes der diagnostischen Validität jedoch nur bedingt (s. Tabelle 11). Die Sensitivität betrug zwar 91%, die Spezifität aber nur 54%. Aufgrund der hohen Rate falsch Positiver wurde ein positiver Vorhersagewert von lediglich 52% erreicht. Die Rate falsch Negativer war dagegen gering, sodass der negative Vorhersagewert 92% betrug. Die Gesamttrefferquote war mit 67% unzureichend.

Der optimale Cut-off-Wert der Skala Affektive Symptome lag bei 15 Punkten. Auch hier war die AUC als zufriedenstellend zu beurteilen ($AUC = .79$, $p < .001$, 95% KI [.72, .85]), Sensitivität (90%) und negativer Vorhersagewert (91%) waren hoch. Die Spezifität (53%), der positive Vorhersagewert (50%) und die Gesamttrefferquote (66%) waren dagegen wegen der hohen Rate falsch Positiver gering (s. Tabelle 11).

Für die Skala Zwang lag der optimale Cut-off-Wert bei 6 Punkten. Die AUC betrug .73 ($p < .001$, 95% KI [.66, .80]). Die Güteindizes der diagnostischen Validität waren dagegen nicht zufriedenstellend (s. Tabelle 11). Die Sensitivität betrug lediglich 54%, die Spezifität 78%. Sowohl der positive Vorhersagewert (57%) als auch der negative Vorhersagewert

(76%) waren gering. Nur 70% der Probanden wurden mit dieser Subskala korrekt kategorisiert.

Tabelle 11

Übereinstimmung der Klassifikation des Antwortverhaltens durch die SN-S-Subskalen mit der Untersuchungsgruppe (Fortsetzung S. 68)

Untersuchungs- gruppe	SN-S-Skala dichotom		Σ
	unauffällig	supernormal	
Soziale Erwünschtheit			
KG (%)	71 (54.2)	60 (45.8)	131 (100)
Fake-KG (%)	6 (8.6)	64 (91.4)	70 (100)
Σ (%)	77 (38.3)	124 (61.7)	201 (100)
Affektive Symptome			
KG (%)	69 (52.7)	62 (47.3)	131 (100)
Fake-KG (%)	7 (10.0)	63 (90.0)	70 (100)
Σ (%)	76 (37.8)	125 (62.2)	201 (100)
Zwang			
KG (%)	102 (77.9)	29 (22.1)	131 (100)
Fake-KG (%)	32 (45.7)	38 (54.3)	70 (100)
Σ (%)	134 (66.7)	67 (33.3)	201 (100)
Dissoziatives Erleben			
KG (%)	71 (54.2)	60 (45.8)	131 (100)
Fake-KG (%)	11 (15.7)	59 (84.3)	70 (100)
Σ (%)	82 (40.8)	119 (59.2)	201 (100)

Fortsetzung Tabelle 11

Untersuchungs- gruppe	SN-S-Skala dichotom		Σ
	unauffällig	supernormal	
Psychotische Symptome			
KG (%)	56 (42.7)	75 (57.3)	131 (100)
Fake-KG (%)	8 (11.4)	62 (88.6)	70 (100)
Σ (%)	64 (31.8)	137 (68.2)	201 (100)
Aggression			
KG (%)	111 (84.7)	20 (15.3)	131 (100)
Fake-KG (%)	24 (34.3)	46 (65.7)	70 (100)
Σ (%)	135 (67.2)	66 (32.8)	201 (100)

Der optimale Cut-off-Wert der Skala Dissoziatives Erleben lag bei 10 Punkten. Die AUC erreichte mit .73 einen zufriedenstellenden Wert ($p < .001$, 95% KI [.66, .80]). Sensitivität (84%) und negativer Vorhersagewert (87%) erreichten akzeptable Werte, Spezifität (54%) und positiver Vorhersagewert (50%) waren dagegen unzureichend. Die Gesamttrefferquote lag lediglich bei 65% (s. Tabelle 11).

Der optimale Cut-off-Wert der Skala Psychotische Symptome lag bei 10 Punkten, die AUC betrug .72. Die Sensitivität (89%) und der negative Vorhersagewert (88%) waren gut. Die Spezifität (43%) und der positive Vorhersagewert (45%) waren so unzureichend, dass lediglich 59% der Probanden korrekt klassifiziert wurden. Denn auch die Mehrheit der KG erreichte auf dieser Skala Werte unterhalb des Cut-offs (s. Tabelle 11).

Für die Skala Aggression betrug der optimale Cut-off-Wert 11 Punkte, die AUC .82 ($p < .001$, 95% KI [.66, .80]). Zwar war die Sensitivität mit 66% gering, Spezifität (85%), positiver Vorhersagewert (70%) und negativer Vorhersagewert (82%) waren aber akzeptabel. Diese Skala hatte die höchste Rate falsch Negativer (s. Tabelle 11), die Gesamttrefferquote betrug daher nur 78%.

5.3.3 Aus der Faktorenanalyse extrahierte Skalen

Der optimale Cut-off-Wert für die Skala Minimierung depressiver Beschwerden lag bei 16 Punkten. Die AUC betrug .80 ($p < .001$, 95% KI [.73, .86]). Die Güteindizes der diagnostischen Validität waren jedoch nicht zufriedenstellend, was auf eine hohe Anzahl falsch Positiver und auch falsch Negativer zurückzuführen war (s. Tabelle 12). Die Sensitivität lag bei 78%, die Spezifität bei 67%, der positive Vorhersagewert bei 56%. Lediglich der negative Vorhersagewert lag mit 85% im akzeptablen Bereich. Insgesamt wurden 71% der Probanden korrekt klassifiziert.

Tabelle 12

Übereinstimmung der Klassifikation des Antwortverhaltens durch die aus der Faktorenanalyse extrahierten SN-S-Subskalen mit der Untersuchungsgruppe

Untersuchungs- gruppe	SN-S-Skala dichotom		Σ
	unauffällig	supernormal	
Minimierung depressiver Beschwerden			
KG (%)	85 (67.5)	41 (32.5)	126 (100)
Fake-KG (%)	15 (22.1)	53 (77.9)	68 (100)
Σ (%)	100 (51.5)	94 (48.5)	194 (100)
Minimierung von Aggression			
KG (%)	76 (58.5)	54 (41.5)	130 (100)
Fake-KG (%)	4 (5.8)	65 (94.2)	69 (100)
Σ (%)	80 (40.2)	119 (59.8)	199 (100)
Minimierung von ungewöhnlichem Erleben			
KG (%)	83 (64.8)	45 (35.2)	128 (100)
Fake-KG (%)	20 (29.0)	49 (71.0)	69 (100)
Σ (%)	103 (52.3)	94 (47.7)	197 (100)

Für die Skala Minimierung von Aggression lag der optimale Cut-off-Wert bei 15 Punkten. Die AUC war mit 84. als gut zu beurteilen ($p < .001$, 95% KI [.78, .89]). Auch Sensitivität

(94%) und negativer Vorhersagewert (95%) erreichten hohe Werte. Dagegen waren Spezifität (58%) und positiver Vorhersagewert (55%) gering. Insgesamt wurde eine Trefferquote von 71% erreicht (s. Tabelle 12).

Die Skala Minimierung von ungewöhnlichem Erleben hatte bei 10 Punkten den optimalen Cut-off-Wert. Die AUC betrug .76 ($p < .001$, 95% KI [.69, .83]). Wegen hoher Raten falsch Positiver und falsch Negativer (s. Tabelle 12) waren die Güteindizes der diagnostischen Validität unzureichend. Die Sensitivität (71%) und der negative Vorhersagewert (81%) erreichten noch etwas bessere Werte als Spezifität (65%) und positiver Vorhersagewert (52%). Die Gesamttrefferquote war gering (67%).

Tabelle 13

Übereinstimmung der Klassifikation des Antwortverhaltens durch die aus der Faktorenanalyse extrahierten SN-S-Subskalen mit der Untersuchungsgruppe in der Kreuzvalidierungsstichprobe

Untersuchungs- gruppe	SN-S-Skala dichotom		Σ
	unauffällig	supernormal	
Minimierung depressiver Beschwerden			
KG (%)	19 (67.9)	9 (32.1)	28 (100)
Fake-KG (%)	10 (52.6)	9 (47.4)	19 (100)
Σ (%)	29 (61.7)	18 (38.3)	47 (100)
Minimierung von Aggression			
KG (%)	16 (57.1)	12 (42.9)	28 (100)
Fake-KG (%)	4 (21.1)	15 (78.9)	19 (100)
Σ (%)	20 (42.6)	27 (57.4)	47 (100)
Minimierung von ungewöhnlichem Erleben			
KG (%)	22 (78.6)	6 (21.4)	28 (100)
Fake-KG (%)	8 (42.1)	11 (57.9)	19 (100)
Σ (%)	30 (63.8)	17 (36.2)	47 (100)

Die Güteindizes wurden i.S. einer Kreuzvalidierung noch einmal mit der KV-Stichprobe berechnet. Hier fiel die Übereinstimmung zwischen Gruppenzugehörigkeit und Klassifikation des Antwortverhaltens noch geringer aus (s. Tabelle 13).

Die Skala Minimierung depressiver Beschwerden hatte in dieser Stichprobe eine Sensitivität von lediglich 47%, weil die Probanden der Fake-KG häufiger unauffällige als auffällige Werte erzielten. Doch auch die Rate falsch Positiver war nicht gering, die Spezifität betrug lediglich 68%. Positiver (50%) und negativer Vorhersagewert (66%) waren ebenso unzureichend wie die Gesamttrefferquote (60%).

Eine etwas bessere Validität hatte die Skala Minimierung von Aggression wegen der geringeren Rate falsch Negativer. Insgesamt waren die Kennwerte aber auch hier unzureichend mit einer Sensitivität von 79%, einer Spezifität von 57% und einem positiven Vorhersagewert von 56%. Lediglich der negative Vorhersagewert erreichte mit 80% einen akzeptablen Wert. Die Gesamttrefferquote war mit 66% gering.

Die höchste Gesamttrefferquote zeigte die Skala Minimierung von ungewöhnlichem Erleben (70%), die jedoch ebenso wie die anderen Güteindizes unzureichend war. Die Sensitivität der Skala betrug 58%, die Spezifität 79%, der positive Vorhersagewert 65% und der negative Vorhersagewert 73%.

5.3.4 PPI-Skalen

Für die Skala Unaufrichtige Beantwortung wurde der im Manual vorgegebene Cut-off-Wert von $T > 60$ für die Kategorisierung des Antwortverhaltens herangezogen.

Zwar zeigte die Skala eine hohe Spezifität (96%), positiver (79%) und negativer Vorhersagewert (72%) lagen im akzeptablen Bereich. Weil auch die Mehrheit der Fake-KG unauffällige Werte erreichte, war die Sensitivität mit 28% unzureichend. Die Gesamttrefferquote betrug 73% (s. Tabelle 14), die AUC lediglich .60 ($p < .05$, 95% KI [.52, .67]).

Der optimale Cut-off-Wert der PPI-SE-Skala lag bei 23 Punkten. Die AUC war mit .76 ($p < .001$, 95% KI [.68, .83]) zwar als zufriedenstellend zu beurteilen. Wegen der hohen Raten an falsch Positiven und falsch Negativen waren aber die Güteindizes der diagnostischen Validität nicht zufriedenstellend. Die Sensitivität betrug 71%, die Spezifität 70%, der positive Vorhersagewert 56% und der negative Vorhersagewert 82%. Es wurden insgesamt nur 70% der Probanden korrekt klassifiziert (s. Tabelle 15).

Tabelle 14

Übereinstimmung der Klassifikation des Antwortverhaltens durch die PPI-UA-Skala mit der Untersuchungsgruppe

Untersuchungs- gruppe	PPI-UA-Skala dichotom		Σ
	unauffällig	auffällig	
KG (%)	124 (96.1)	5 (3.9)	129 (100)
Fake-KG (%)	49 (72.1)	19 (27.9)	68 (100)
Σ (%)	173 (87.8)	24 (12.2)	197 (100)

Tabelle 15

Übereinstimmung der Klassifikation des Antwortverhaltens durch die PPI-SE-Skala mit der Untersuchungsgruppe

Untersuchungs- gruppe	PPI-SE-Skala dichotom		Σ
	unauffällig	auffällig	
KG (%)	85 (69.7)	37 (30.3)	122 (100)
Fake-KG (%)	19 (28.8)	47 (71.2)	66 (100)
Σ (%)	104 (55.3)	84 (44.7)	188 (100)

5.4 Konvergente Validität

5.4.1 Dimensionale Übereinstimmung mit anderen Kontrollskalen

Die Korrelationen der SN-S-Skalen mit den anderen Kontrollskalen zur Erfassung von positiven Antwortverzerrungen sind in Tabelle 16 dargestellt, die Streudiagramme für den SN-S-Gesamtwert können dem Anhang entnommen werden (Anhang C, Abbildung 2 und 3). Der Gesamtwert korrelierte signifikant mit den beiden PPI-Skalen, wobei sich ein etwas höherer Zusammenhang mit der SE-Skala zeigte, $r(228) = .36$, während die Korrelation mit der UA-Skala eher gering ausfiel, $r(238) = .16$. Mit den MMPI-Kontrollskalen zeigte sich eine hohe Übereinstimmung ($r = .60-.80$). Auch mit der ESI-FR-, $r(9) = .49$, und der 16PF-IM, $r(10) = -.47$, korrelierte der Gesamtwert moderat bis hoch, das Signifikanzni-

veau wurde aber nicht erreicht. Ein eher geringer Zusammenhang bestand mit der FAF-Off-Skala, $r(21) = .24$.

Tabelle 16

Korrelationen des SN-S-Gesamtwerts (GW) und der Subskalen mit den anderen Kontrollskalen zur Erfassung positiver und negativer Antwortverzerrungen

Kontrollskalen	SN-S									
	GW	SE	AS	ZW	DE	PS	AGG	MDe	MA	MUE
PPI-UA	-.16 [*]	-.30 ^{**}	-.11	-.13 [*]	-.04	-.01	-.24 ^{**}	-.10	-.27 ^{**}	-.05
PPI-SE	-.36 ^{**}	-.39 ^{**}	-.31 ^{**}	-.28 ^{**}	-.15 [*]	-.19 ^{**}	-.38 ^{**}	-.30 ^{**}	-.41 ^{**}	-.20 ^{**}
FAF-Off	.24	.22	.27	.20	.14	.23	.12	.24	.13	.27
16PF-IM	-.47	-.57	-.48	-.45	-.30	-.49	-.46	-.38	-.46	-.37
ESI-FR	.49	.63	.53	.24	.15	-.12	.60	.36	.60	.35
MMPI L	-.60	-.63 [*]	.65 [*]	.46	-.43	-.14	-.71 [*]	-.74 ^{**}	-.68 [*]	-.38
MMPI F	.79 ^{**}	.65 [*]	.76 ^{**}	.56	.70 [*]	.30	.72 [*]	.75 ^{**}	.86 ^{**}	.59
MMPI K	-.80 ^{**}	-.36	-.78 ^{**}	-.52	-.91 ^{**}	-.62 [*]	-.70 [*]	-.76 ^{**}	-.72 [*]	-.84 ^{**}

Anmerkung: ^{*} $p < .05$, ^{**} $p < .01$

Bei den SN-S-Subskalen zeigte sich überwiegend ein ähnliches Muster. Überwiegend moderate bis hohe Zusammenhänge zeigten sich mit den MMPI-Skalen ($r = .14-.91$). Die 16PF-IM-Skala hatte die höchste Korrelation mit der Skala soziale Erwünschtheit, $r(10) = .57$, die übrigen Korrelationen lagen überwiegend im moderaten Bereich ($r = .30-.48$), erreichten aber keine Signifikanz. Die Skalen, die eine Minimierung aggressiver Symptome (AGG, MA) erfassen, zeigten die größten Zusammenhänge mit der MMPI-F-Skala ($r = .70-.72$) und nicht mit der Offenheit-Skala des FAF ($r = .12-.13$). Mit dieser fielen alle Korrelationen eher gering aus ($r = .12-.27$). Die ESI-FR-Skala zeigte den größten Zusammenhang mit den SN-S-Skalen Soziale Erwünschtheit, $r(9) = .63$, Aggression und Minimierung von Aggression, jeweils $r(9) = .60$. Mit der Skala Psychotische Symptome zeigte sie dagegen keine Übereinstimmung, $r(9) = .12$. Einen moderaten, jedoch nicht signifikanten

ten Zusammenhang zeigte sie aber mit der Skala, die eine Minimierung von depressiven Symptomen erfasst, $r(9) = .53$.

5.4.2 Kategoriale Übereinstimmung mit anderen Kontrollskalen

Die Klassifikation des Antwortverhaltens durch den SN-S-Gesamtwert stimmte lediglich bei 51% der Probanden mit der Klassifikation der PPI-UA-Skala überein, $\chi^2(1, N = 238) = .18, p = .745$ (s. Tabelle 17). Insbesondere der Anteil von Probanden, die in der SN-S auffällige Werte, auf der UA-Skala aber unauffällige Werte erzielt hatten (40%), war hoch.

Mit der SE-Skala zeigte die SN-S-Klassifikation dagegen eine signifikante Übereinstimmung, $\chi^2(1, N = 228) = 13.85, p < .001$ (s. Tabelle 17). Das Antwortverhalten von 62% der Probanden wurde übereinstimmend als auffällig oder unauffällig klassifiziert.

Mit den Skalen FAF-Off, 16PF-IM und ESI-FR zeigte die SN-S keine signifikante Übereinstimmung in der Klassifikation (s. Tabelle 17). Das Antwortverhalten von 71%, 78% bzw. 56% der Probanden wurde jeweils übereinstimmend als auffällig oder unauffällig klassifiziert. Eine genauere Betrachtung der Kreuztabellen zeigte, dass es nur einen (FAF-Off) bzw. keine Probanden (16PF-IM, ESI-FR) gab, die auf diesen Kontrollskalen auffällige, aber in der SN-S unauffällige Werte erzielten. Höher war dagegen der Anteil von Probanden, die in den anderen Kontrollskalen unauffällige Werte, aber in der SN-S auffällige Werte erzielten (FAF-Off: 23%, 16PF-IM: 22%; ESI-FR: 33%).

Tabelle 17

Übereinstimmung der Klassifikation des Antwortverhaltens durch den SN-S-Gesamtwert und andere Kontrollskalen zur Erfassung positiver Antwortverzerrungen (Fortsetzung S. 75)

Kontrollskala	Antwortverhalten	SN-S-Gesamtwert dichotom		Σ
		unauffällig	supernormal	
PPI-UA ^a	unauffällig	96	95	191
	auffällig	22	25	47
	Σ	118	120	238

Fortsetzung Tabelle 17		SN-S-Gesamtwert dichotom		
Kontrollskala	Antwortverhalten	unauffällig	supernormal	Σ
PPI-SE ^b	unauffällig	70	46	116
	auffällig	40	72	112
	Σ	110	118	228
FAF-Off ^c	unauffällig	12	5	17
	auffällig	1	3	4
	Σ	13	8	21
16PF-IM ^d	unauffällig	4	2	6
	auffällig	0	3	3
	Σ	4	5	9
ESI-FR ^e	unauffällig	2	3	5
	auffällig	1	3	4
	Σ	3	6	9

Anmerkung: ^a $\chi^2(1, N = 238) = .18, p = .745$ ^b $\chi^2(1, N = 228) = 13.85, p < .001$; ^c $p = .253$; ^d $p = .167$; ^e $p = 1.0$

Mit den MMPI-Skalen ergaben sich ebenso keine signifikanten Übereinstimmungen (s. Tabelle 18). Nur 25% (L-Skala) bzw. 50% (F- und K-Skala) wurden jeweils übereinstimmend klassifiziert. Die Probanden erzielten deutlich häufiger in der SN-S auffällige Werte als auf L-, F- oder K-Skala. Dagegen wurde das Antwortverhalten von keinem Probanden mit einem unauffälligen SN-S-Wert auf einer der drei MMPI-Skalen als dissimulierend eingestuft.

Auch mit den SN-S-Subskalen stimmten nur wenige Kontrollskalen bzw. lediglich die Kontrollskalen des PPI-R signifikant überein. Die Klassifikation der PPI-UA-Skala zeigte eine signifikante Übereinstimmung mit den Skalen Zwang, $\chi^2(1, N = 238) = 4.07, p < .05$, und Minimierung von Aggression, $\chi^2(1, N = 236) = 6.18, p < .05$ (s. Anhang C, Tabelle 28). Die Klassifikation der PPI-SE-Skala zeigte signifikante Zusammenhänge mit den Skalen Soziale Erwünschtheit, $\chi^2(1, N = 228) = 18.80, p < .001$, Affektive Symptome, $\chi^2(1, N = 228) = 10.84, p < .01$, Zwang, $\chi^2(1, N = 228) = 5.93, p < .05$, Psychotische Symptome, χ^2

(1, $N = 228$) = 6.45, $p < .05$, Aggression, χ^2 (1, $N = 228$) = 22.19, $p < .001$, Minimierung depressiver Beschwerden, χ^2 (1, $N = 221$) = 13.72, $p < .001$, Minimierung von Aggression, χ^2 (1, $N = 226$) = 24.26, $p < .001$, sowie Minimierung von ungewöhnlichem Erleben, χ^2 (1, $N = 224$) = 6.46, $p < .05$. Die Kreuztabellen können dem Anhang entnommen werden (Anhang C, Tabelle 28).

Tabelle 18

Übereinstimmung der Klassifikation des Antwortverhaltens durch den SN-S-Gesamtwert und die MMPI-Kontrollskalen

Kontrollskala	Antwortverhalten	SN-S-Gesamtwert dichotom		Σ
		unauffällig	supernormal	
MMPI-L ^a	mögl. Simulation	0	1	1
	gültiges Profil	1	2	3
	Abwehr	1	2	3
	Dissimulation	0	1	1
	Σ	2	6	8
MMPI-F ^b	gültiges Profil	2	4	6
	mögl. Dissimulation	0	2	2
	Σ	2	6	8
MMPI-K ^c	Selbstöffnung	2	4	6
	Dissimulation	0	2	2
	Σ	2	6	8

Anmerkung: ^a $p = 1.0$; ^b $p = 1.0$; ^c $p = 1.0$

5.5 Erfassung von negativen Antwortverzerrungen mit der SN-S

Der SN-S-Gesamtwert und alle Subskalen wiesen signifikante moderate bis hohe Korrelationen mit dem SFSS-Gesamtwert auf ($r = .41-.81$; s. Tabelle 19). Die Streudiagramme können dem Anhang entnommen werden (Anhang C, Abbildung 4 und 5). Auch mit den SFSS-Subskalen Affektive Symptome, Amnestische Störung und Neurologische Beeinträchtigung ergaben sich signifikante moderate bis hohe Zusammenhänge ($r = .34-.84$). Die

Skala Niedrige Intelligenz korrelierte nicht mit den Skalen Soziale Erwünschtheit, Dissoziatives Erleben, Minimierung von Aggression und Minimierung von ungewöhnlichem Erleben, mit den übrigen SN-S-Skalen fielen die Korrelationen eher gering bis moderat aus ($r = .24-.39$). Die SFSS-Skala Psychose zeigte mit allen SN-S-Skalen signifikante Zusammenhänge, mit den Skalen Aggression und Minimierung von Aggression fielen die Korrelationen aber gering aus ($r = .26-.29$).

Tabelle 19

Korrelationen des SN-S-Gesamtwerts (GW) und der SN-S-Subskalen mit SFSS-Gesamtwert (GW) und den SFSS-Subskalen

SFSS	SN-S									
	GW	SE	AS	ZW	DE	PS	AGG	MDe	MA	MUE
SFSS-GW	.78**	.41**	.77**	.71**	.64**	.73**	.47**	.81**	.47**	.64**
SFSS-AS	.79**	.44**	.79**	.70**	.61**	.73**	.46**	.84**	.46**	.65**
SFSS-Am	.65**	.30**	.65**	.61**	.53**	.56**	.48**	.67**	.47**	.49**
SFSS-N	.68**	.34**	.69**	.64**	.57**	.63**	.39**	.74**	.40**	.55**
SFSS-NI	.24*	.04	.22*	.24*	.17	.25*	.39**	.29**	.16	.08
SFSS-P	.59**	.40**	.53**	.51**	.56**	.63**	.26*	.52**	.29**	.59**

Anmerkung: * $p < .05$, ** $p < .01$

Die ROC-Analyse für den SN-S-Gesamtwert mit dem SFSS-Ergebnis als Außenkriterium erbrachte einen Cut-off-Wert für das Vorliegen negativer Antwortverzerrungen von 78 Punkten. Die AUC war hoch ($AUC = .91$, $p < .001$, 95% KI [.85, .97]). In der SN-S erzielten 90% der Probanden mit unauffälligem SFSS-Gesamtwert einen Wert unterhalb des Cut-offs von 78 Punkten (Spezifität). Allerdings wurde nur bei 78% der Probanden mit auffälligem SFSS-Gesamtwert das Antwortverhalten durch die SN-S als aggravierend oder simulierend eingestuft (Sensitivität). Dementsprechend wurde zwar ein hoher negativer Vorhersagewert (93%) erreicht, der positive Vorhersagewert war dagegen eher gering (69%). Insgesamt wurde das Antwortverhalten von 87% der Probanden übereinstimmend als unauffällig oder auffällig klassifiziert (s. Tabelle 20).

Tabelle 20

Übereinstimmung der Klassifikation des Antwortverhaltens durch den SN-S-Gesamtwert und den SFSS-Gesamtwert

		SN-S-Gesamtwert dichotom		Σ
		unauffällig	aggravierend	
SFSS-Gesamtwert dichotom	unauffällig (%)	71 (89.9)	8 (10.1)	79 (100)
	aggravierend (%)	5 (21.7)	18 (78.3)	23 (100)
	Σ (%)	76 (74.5)	26 (25.5)	102 (100)

5.6 Die Häufigkeit des Auftretens von positiven Antwortverzerrungen

5.6.1 Mittelwertsunterschiede zwischen den verschiedenen Substichproben

Die deskriptiven Statistiken der SN-S in den Substichproben können dem Anhang entnommen werden (Anhang C, Tabelle 29), die Verteilung des SN-S-Gesamtwert der Abbildung 1, die Verteilung der Skalenwerte dem Anhang (Anhang C, Abbildung 6-8).

Die einzelnen Untersuchungsgruppen erzielten signifikant unterschiedliche SN-S-Gesamtwerte, $F(5, 303) = 21.77, p < .001$. Die Bonferroni Post hoc Tests ($p < .05$) zeigten, dass Schuldfähigkeitsprobanden ($M = 71.5, SD = 21.3$) signifikant höhere Werte erzielten als Prognose- ($M = 60.1, SD = 13.3$), Gesundheitsnachweis- ($M = 52.1, SD = 10.4$) und Fake-KG-Probanden ($M = 51.8, SD = 7.1$). Die Prognose-, Gesundheitsnachweis-, KG- ($M = 66.8, SD = 11.9$) und Fake-KG-Probanden erzielten jeweils signifikant niedrigere Werte als die Zahlungsbegehren-Probanden ($M = 77.6, SD = 19.8$). Die Gesundheitsnachweisprobanden unterschieden sich darüber hinaus auch von den KG-Probanden signifikant, nicht aber von den Fake-KG-Probanden, welche jedoch signifikant niedrigere Werte erzielten als die übrigen Substichproben.

Auch auf allen SN-S-Subskalen wurden von den Stichproben signifikant unterschiedliche Werte erreicht. Auf der Skala Soziale Erwünschtheit, $F(5, 303) = 13.01, p < .001$, erzielten Schuldfähigkeits- ($M = 11.0, SD = 2.6$), Prognose- ($M = 10.7, SD = 1.9$), Zahlungsbegehren- ($M = 10.9, SD = 2.6$) und KG-Probanden ($M = 10.6, SD = 2.4$), nicht aber die Gesundheitsnachweisprobanden ($M = 8.5, SD = 1.8$), jeweils signifikant höhere Werte als die Probanden der Fake-KG ($M = 8.3, SD = 1.7$).

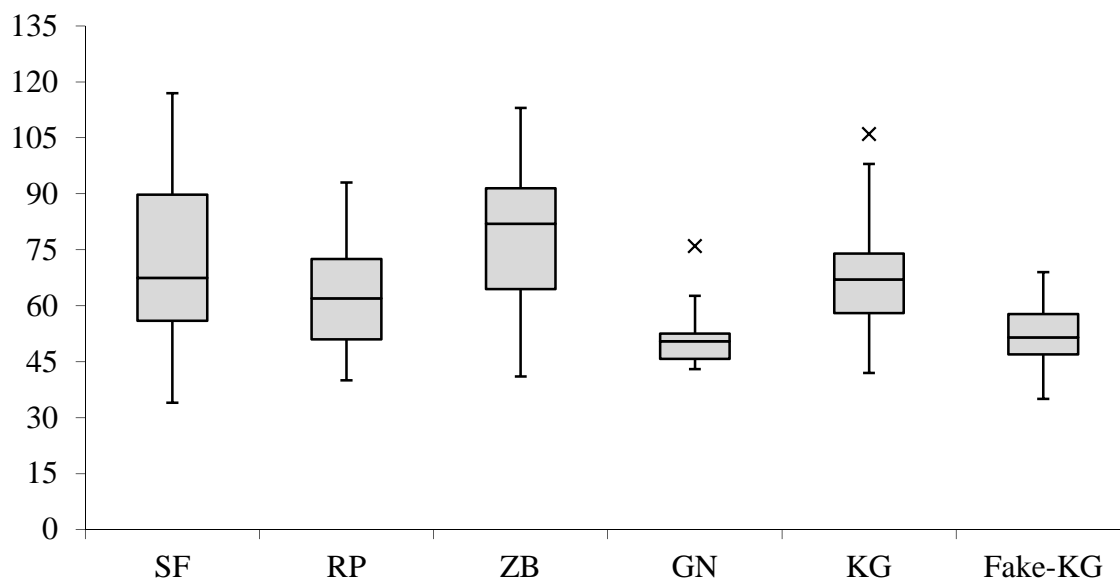


Abbildung 1. Verteilung des SN-S-Gesamtwerts in den Substichproben

Auf der Skala Affektive Symptome, $F(5, 303) = 19.00, p < .001$, erzielten Schuldfähigkeits- ($M = 16.7, SD = 5.5$) und Zahlungsbegehrenprobanden ($M = 19.5, SD = 5.5$) jeweils signifikant höhere Werte als Prognose- ($M = 13.5, SD = 4.6$), Gesundheitsnachweis- ($M = 11.6, SD = 3.0$) und Fake-KG-Probanden ($M = 11.4, SD = 2.6$). Die Zahlungsbegehrenprobanden erreichten außerdem signifikant höhere Werte als die KG-Probanden ($M = 15.2, SD = 3.9$). Diese wiederum hatten signifikant höhere Werte als die Fake-KG-Probanden.

Auf der Skala Zwang, $F(5, 303) = 12.94, p < .001$, hatten die Schuldfähigkeits- ($M = 7.9, SD = 3.1$) und Zahlungsbegehrenprobanden ($M = 8.6, SD = 2.9$) jeweils signifikant höhere Werte als die Prognose- ($M = 5.9, SD = 2.0$) und Fake-KG-Probanden ($M = 5.4, SD = 1.3$). Die Zahlungsbegehrenprobanden erreichten außerdem signifikant höhere Werte als die KG-Probanden ($M = 7.0, SD = 2.1$). Diese hatten wiederum signifikant höhere Werte als Prognose-, Gesundheitsnachweis- ($M = 6.0, SD = 2.0$) und Fake-KG-Probanden.

Auf der Skala Psychotische Symptome, $F(5, 303) = 7.63, p < .001$, hatten Schuldfähigkeits- ($M = 9.9, SD = 4.7$), Zahlungsbegehren- ($M = 10.4, SD = 3.9$) und KG-Probanden ($M = 9.4, SD = 2.6$) jeweils signifikant höhere Werte als Prognose- ($M = 7.7, SD = 2.3$) und Fake-KG-Probanden ($M = 7.6, SD = 1.6$). Die Gesundheitsnachweisgruppe ($M = 7.5, SD = 1.9$) unterschied sich von keiner der anderen Stichproben signifikant.

Auf der Skala Dissoziatives Erleben, $F(5, 303) = 7.92, p < .001$, hatten Schuldfähigkeits- ($M = 10.1, SD = 3.8$), Zahlungsbegehren- ($M = 11.0, SD = 3.9$) und KG-Probanden ($M =$

9.7, $SD = 2.5$) jeweils signifikant höhere Werte als Prognose- ($M = 9.2$, $SD = 2.2$), Gesundheitsnachweis- ($M = 8.4$, $SD = 2.6$) und Fake-KG-Probanden ($M = 7.8$, $SD = 1.8$).

Auf der Skala Aggression, $F(5, 303) = 15.69$, $p < .001$, erreichten Schuldfähigkeits- ($M = 13.9$, $SD = 5.0$) und Zahlungsbegehrenprobanden ($M = 14.4$, $SD = 4.6$) jeweils höhere Werte als Prognose- ($M = 11.5$, $SD = 3.4$), Gesundheitsnachweis- ($M = 9.1$, $SD = 1.6$) und Fake-KG-Probanden ($M = 9.8$, $SD = 2.1$). Die KG ($M = 13.2$, $SD = 3.0$) hatte signifikant höhere Werte als die Gesundheitsnachweisgruppe und die Fake-KG.

Auf der Skala Minimierung von depressiven Beschwerden, $F(5, 295) = 21.69$, $p < .001$, hatte die Gruppe Zahlungsbegehren ($M = 25.4$, $SD = 7.1$) signifikant höhere Werte als alle anderen Stichproben. Die Schuldfähigkeitsprobanden ($M = 20.3$, $SD = 7.2$) erzielten signifikant höhere Werte als Prognose- ($M = 16.7$, $SD = 5.6$), Gesundheitsnachweis- ($M = 14.1$, $SD = 3.8$) und Fake-KG-Probanden ($M = 13.6$, $SD = 3.0$). Auch Prognose- und KG-Probanden ($M = 18.6$, $SD = 5.1$), nicht aber die Gesundheitsnachweisprobanden, hatten jeweils signifikant höhere Werte als die Fake-KG.

Auf der Skala Minimierung von Aggression, $F(5, 301) = 15.80$, $p < .001$, hatten Schuldfähigkeits- ($M = 15.9$, $SD = 5.7$), Zahlungsbegehren- ($M = 16.7$, $SD = 5.4$) und KG-Probanden ($M = 15.4$, $SD = 3.6$) jeweils signifikant höhere Werte als Gesundheitsnachweis- ($M = 10.8$, $SD = 1.5$) und Fake-KG-Probanden ($M = 11.2$, $SD = 2.2$). Die Zahlungsbegehren- und KG-Probanden erreichten außerdem jeweils signifikant höhere Werte als die Prognoseprobanden ($M = 13.3$, $SD = 4.1$).

Auf der Skala Minimierung von ungewöhnlichem Erleben, $F(5, 297) = 9.60$, $p < .001$, erzielten Schuldfähigkeits- ($M = 12.2$, $SD = 4.0$) und Zahlungsbegehrenprobanden ($M = 12.2$, $SD = 4.7$) jeweils signifikant höhere Werte als Prognose- ($M = 9.7$, $SD = 2.8$) und Fake-KG-Probanden ($M = 8.6$, $SD = 2.0$). Auch die KG ($M = 11.1$, $SD = 3.2$) hatte signifikant höhere Werte als die Fake-KG. Die Gruppe Gesundheitsnachweis ($M = 9.4$, $SD = 3.3$) unterschied sich von keiner anderen Substichprobe signifikant.

5.6.2 Häufigkeitsunterschiede zwischen den verschiedenen Substichproben

Die Substichproben unterschieden sich signifikant hinsichtlich der Häufigkeit supernormalen Antwortverhaltens, $\chi^2(5, N = 309) = 86.99$, $p < .001$ (s. Tabelle 21). Am häufigsten erzielten die Fake-KG- (94%) und die Gesundheitsnachweisprobanden (88%) auffällige SN-S-Gesamtwerte. Auch etwas mehr als die Hälfte der Prognoseprobanden (51%) zeigte supernormales Antwortverhalten, aber nur 38% der Schuldfähigkeitsprobanden und 31% der KG. Am seltensten antworteten die Zahlungsbegehrenprobanden supernormal (22%).

Nach Bonferroni-Korrektur zeigten die Post hoc Tests lediglich signifikante Gruppenunterschiede jeweils zwischen der Fake-KG und den Gruppen Schuldfähigkeit, Prognose, Zahlungsbegehren und KG, außerdem zwischen den Gruppen Gesundheitsnachweis und Zahlungsbegehren bzw. KG.

Auch sozial erwünschtes Antwortverhalten trat in den Stichproben unterschiedlich oft auf, $\chi^2(5, N = 309) = 60.07, p < .001$ (s. Tabelle 21). Über 90% der Fake-KG- und 88% der Gesundheitsnachweisprobanden antworteten sozial erwünscht. In der KG erreichten 46% auffällige Werte, in der Gruppe Schuldfähigkeit 38%, in der Gruppe Prognose 37% und in der Gruppe Zahlungsbegehren 30%. Signifikante Gruppenunterschiede zeigten sich aber auch hier nur zwischen Fake-KG und den anderen Gruppen mit Ausnahme der Gesundheitsnachweisprobanden.

Affektive Symptome wurden von den Stichproben unterschiedlich häufig minimiert, $\chi^2(5, N = 309) = 58.01, p < .001$ (s. Tabelle 21). Am häufigsten erzielten erneut Fake-KG (90%) und Gesundheitsnachweisprobanden (88%) auffällige Werte. Auch 62% der Prognoseprobanden leugneten das Vorliegen affektiver Symptome, aber nur 47% der KG und 41% der Schuldfähigkeitsprobanden. In der Gruppe Zahlungsbegehren erreichten 17% auffällige Skalenwerte. Post hoc unterschied sich erneut die Fake-KG von den Gruppen Schuldfähigkeit, Prognose, Zahlungsbegehren und KG. Die Zahlungsbegehrenprobanden unterschieden sich außerdem jeweils signifikant von Prognose-, Gesundheitsnachweis- und KG-Probanden.

Auf der Skala Zwang zeigten sich ebenfalls signifikante Gruppenunterschiede, $\chi^2(5, N = 309) = 31.15, p < .001$ (s. Tabelle 21). Hier wurden insgesamt am seltensten auffällige Werte erzielt: von 54% der Fake-KG, 51% der Prognose- und 50% der Gesundheitsnachweisprobanden. In der Gruppe Schuldfähigkeit wurden von 27% der Probanden Zwangssymptome minimiert, in der KG von 22% der Probanden und in der Gruppe Zahlungsbegehren von 17%. Fake-KG-, Prognose- sowie Gesundheitsnachweisprobanden dissimulierten hier jeweils häufiger als Zahlungsbegehrenprobanden und KG-Probanden.

Tabelle 21

Häufigkeit auffälliger Werte auf den einzelnen SN-S-Skalen in den verschiedenen Untersuchungsgruppen (Fortsetzung S. 83)

	SF	RP	ZB	GN	KG	Fake-KG
SN-S-Skala dichotom	(n = 34)	(n = 43)	(n = 23)	(n = 8)	(n = 131)	(n = 70)
Gesamtwert						
unauffällig (%)	21 (61.8)	21 (48.8)	18 (78.3)	1 (12.5)	90 (68.7)	4 (5.7)
supernormal (%)	13 (38.2)	22 (51.2)	5 (21.7)	7 (87.5)	41 (31.3)	66 (94.3)
Soziale Erwünschtheit						
unauffällig (%)	21 (61.8)	27 (62.8)	16 (69.9)	1 (12.5)	71 (54.2)	6 (8.6)
supernormal (%)	13 (38.2)	16 (37.2)	7 (30.4)	7 (87.5)	60 (45.8)	64 (91.4)
Affektive Symptome						
unauffällig (%)	20 (58.8)	16 (37.2)	19 (82.6)	1 (12.5)	69 (52.7)	7 (10.0)
supernormal (%)	14 (41.2)	27 (62.8)	4 (17.4)	7 (87.5)	62 (47.3)	63 (90.0)
Zwang						
unauffällig (%)	25 (73.5)	21 (48.8)	19 (82.6)	4 (50.0)	102 (77.9)	32 (45.7)
supernormal (%)	9 (26.5)	22 (51.2)	4 (17.4)	4 (50.0)	29 (22.1)	38 (54.3)
Dissoziatives Erleben						
unauffällig (%)	18 (52.9)	17 (39.5)	15 (65.2)	2 (25.0)	71 (54.2)	11 (15.7)
supernormal (%)	16 (47.1)	26 (60.5)	8 (34.8)	6 (75.0)	60 (45.8)	59 (84.3)

Fortsetzung Tabelle 21	SF (n = 34)	RP (n = 43)	ZB (n = 23)	GN (n = 8)	KG (n = 131)	Fake-KG (n = 70)
Psychotische Symptome						
unauffällig (%)	13 (38.2)	7 (16.3)	12 (52.2)	1 (12.5)	56 (42.7)	8 (11.4)
supernormal (%)	21 (61.8)	36 (83.7)	11 (47.8)	7 (87.5)	75 (57.3)	62 (88.6)
Aggression						
unauffällig (%)	24 (70.6)	27 (62.8)	17 (73.9)	1 (12.5)	111 (84.7)	24 (34.3)
supernormal (%)	10 (29.4)	16 (37.2)	6 (26.1)	7 (87.5)	20 (15.3)	46 (65.7)
Minimierung depressiver Beschwerden						
unauffällig (%)	24 (72.7)	22 (51.2)	21 (91.3)	2 (25.0)	85 (67.5)	15 (22.1)
supernormal (%)	9 (27.3)	21 (48.8)	2 (8.7)	6 (75.0)	41 (32.5)	53 (77.9)
Minimierung von Aggression						
unauffällig (%)	16 (47.1)	16 (37.2)	16 (69.6)	0	76 (58.8)	4 (5.8)
supernormal (%)	18 (52.9)	27 (62.8)	7 (30.4)	8 (100)	54 (41.4)	65 (94.2)
Minimierung von ungewöhnlichem Erleben						
unauffällig (%)	21 (61.8)	20 (46.5)	13 (59.1)	3 (42.9)	83 (64.8)	20 (29.0)
supernormal (%)	13 (38.2)	23 (53.5)	9 (40.9)	4 (57.1)	45 (35.2)	49 (71.0)

Dissoziatives Erleben wurde von den Untersuchungsgruppen auch unterschiedlich oft minimiert, $\chi^2(5, N = 309) = 35.15, p < .001$ (s. Tabelle 21). Am häufigsten erzielten Fake-KG- (84%), Gesundheitsnachweis- (75%) und Prognoseprobanden (61%) auffällige Werte. Auch 47% der Schuldfähigkeitsprobanden, 46% der KG und 35% der Zahlungsbegehrenprobanden unterschritten den Cut-off-Wert. Die Post hoc Tests zeigten signifikante Unterschiede zwischen Fake-KG und den Gruppen Schuldfähigkeit, Zahlungsbegehren, KG und Prognose.

Psychotische Symptome wurden insgesamt recht häufig, aber von den Substichproben unterschiedlich oft verneint, $\chi^2(5, N = 309) = 32.04, p < .001$ (s. Tabelle 21). Jeweils über 80% der Fake-KG-, Prognose- und Gesundheitsnachweisprobanden erzielten auffällige Werte, außerdem 62% der Schuldfähigkeitsprobanden, 57% der KG-Probanden und 48% der Zahlungsbegehrenprobanden. In den Post hoc Tests unterschieden sich Prognose- und Fake-KG-Probanden jeweils signifikant von KG- und Zahlungsbegehrenprobanden. Die Fake-KG erzielte außerdem häufiger als die Gruppe Schuldfähigkeit auffällige Werte.

Aggression wurde von den Untersuchungsgruppen unterschiedlich oft dissimuliert, $\chi^2(5, N = 309) = 63.24, p < .001$ (s. Tabelle 21). Am häufigsten erzielten Gesundheitsnachweis- (88%) und Fake-KG-Probanden (66%) auffällige Werte, seltener Prognose- (37%) und Schuldfähigkeitsprobanden (29%). In der Gruppe Zahlungsbegehren erzielten 26% Werte unterhalb des Cut-off-Werts, in der KG lediglich 15%. Bei dieser Skala zeigten die Post hoc Tests die meisten signifikanten Unterschiede. Die Gesundheitsnachweis- und Fake-KG-Probanden unterschieden sich jeweils signifikant von den KG-, Zahlungsbegehren-, Schuldfähigkeits- und Prognoseprobanden.

Auf der Skala Minimierung von depressiven Beschwerden wurden von den Gruppen unterschiedlich oft auffällige Werte erzielt, $\chi^2(5, N = 301) = 57.46, p < .001$ (s. Tabelle 21), am häufigsten von den Fake-KG- (78%) und den Gesundheitsnachweisprobanden (75%). Seltener minimierten Prognose- (49%), KG- und Schuldfähigkeitsprobanden (27%) solche Beschwerden. Von den Zahlungsbegehrenprobanden erzielten nur knapp 9% Werte unterhalb des Cut-offs. Die Post hoc Tests zeigten signifikante Unterschiede zwischen Fake-KG und den Gruppen Schuldfähigkeit, Zahlungsbegehren und KG. Gesundheitsnachweis- und Prognoseprobanden unterschieden sich signifikant von den Zahlungsbegehrenprobanden.

Signifikante Gruppenunterschiede zeigten sich auch auf der Skala Minimierung von Aggression, $\chi^2(5, N = 307) = 65.44, p < .001$ (s. Tabelle 21). Alle Gesundheitsnachweisprobanden und 94% der Fake-KG erzielten hier auffällige Werte. Auch 63% der Prognose- und 53% der Schuldfähigkeitsprobanden erreichten Werte unterhalb des Cut-offs, aber nur 41% der KG und 30% der Zahlungsbegehrenprobanden. Post hoc zeigten sich signifikante Unterschiede zwischen Fake-KG und den anderen Gruppen mit Ausnahme der Gesundheitsnachweisgruppe. Diese Probanden unterschieden sich signifikant von den KG- und den Zahlungsbegehrenprobanden.

Auf der Skala Minimierung von ungewöhnlichem Erleben zeigten sich zwar signifikante, insgesamt aber die wenigsten Gruppenunterschiede, $\chi^2(5, N = 303) = 25.56, p < .001$ (s. Tabelle 21). In der Fake-KG erzielten 71% Werte unterhalb des Cut-offs, in der Gruppe

Gesundheitsnachweis 57%, in der Gruppe Prognose 54% und in der Gruppe Zahlungsbegehren immerhin 41%. Nur wenig seltener erzielten Schuldfähigkeits- (38%) und KG-Probanden (35%) auffällige Werte. Post hoc war lediglich der Unterschied zwischen Fake-KG und KG bzw. der Gruppe Schuldfähigkeit signifikant.

5.7 Auswirkung von positiven Antwortverzerrungen auf testpsychologische Befunde

5.7.1 Aggressionsbereitschaft

Zwischen FAF-Agg und PFT-E zeigte sich eine hohe signifikante Korrelation, $r(17) = .69$, $p < .01$. Der FHT-AOS korrelierte moderat negativ, aber nicht signifikant mit dem FAF-Agg, $r(16) = -.45$, $p = .079$, und nicht mit dem PFT-E, $r(33) = .12$, $p = .520$.

Der SN-S-Gesamtwert korrelierte signifikant und hoch mit dem FAF-Agg, $r(21) = .66$, $p < .01$, und moderat mit dem PFT-E, $r(55) = .42$, $p < .01$, aber nicht signifikant mit dem FHT-AOS, $r(35) = .15$, $p = .395$. Die Subskala Aggression zeigte ebenfalls einen signifikanten hohen Zusammenhang mit dem FAF-Agg, $r(21) = .72$, $p < .001$, einen signifikanten moderaten Zusammenhang mit dem PFT-E, $r(55) = .38$, $p < .01$, und keinen Zusammenhang mit dem FHT-AOS, $r(35) = .04$, $p = .821$. Auch die Skala Minimierung von Aggression korrelierte hoch mit dem FAF-Agg, $r(21) = .74$, $p < .001$, moderat mit dem PFT-E, $r(55) = .39$, $p < .01$, und nicht mit dem FHT-AOS, $r(35) = .05$, $p = .78$.

Die Probanden mit supernormalem Antwortverhalten berichteten im FAF eine signifikant geringere Aggressionsbereitschaft ($M = 38.1$, $SD = 7.1$) als die Probanden mit unauffälligem Antwortverhalten ($M = 51.5$, $SD = 9.0$), $t(19) = 3.59$, $p < .01$. Der FHT-AOS, $t(33) = 1.76$, $p = .088$, und die Anzahl der E-Antworten im PFT, $t(53) = 1.71$, $p = .093$, unterschieden sich nicht signifikant zwischen Probanden mit supernormalem und Probanden mit unauffälligem Antwortverhalten (s. Tabelle 22).

Die Probanden, die auf der Subskala Aggression auffällige Werte erzielten, berichteten im FAF eine geringere Aggressionsbereitschaft ($M = 38.4$, $SD = 4.8$) als die Probanden mit unauffälligen Werten ($M = 49.0$, $SD = 10.7$), $t(19) = 2.13$, $p < .05$. Der FHT-AOS, $t(33) = .23$, $p = .818$, und die Anzahl der E-Antworten im PFT, $t(53) = 1.86$, $p = .069$, unterschieden sich nicht signifikant zwischen den beiden Gruppen (s. Tabelle 22).

Die Probanden, die auf der Skala Minimierung von Aggression Werte unterhalb des Cut-offs erzielten, berichteten im FAF eine signifikant geringere Aggressionsbereitschaft ($M = 39.6$, $SD = 9.3$) als Probanden mit Werten oberhalb des Cut-offs ($M = 54.0$, $SD = 5.6$), t

(19) = 4.23, $p < .001$. Außerdem gaben die Probanden mit Minimierung von Aggression signifikant weniger E-Antworten im PFT ($M = 1.8$, $SD = 1.8$) als die Probanden ohne ein solches Antwortverhalten ($M = 3.1$, $SD = 2.2$), $t(53) = 2.38$, $p < .05$. Im FHT zeigten sich keine signifikanten Gruppenunterschiede, $t(33) = .55$, $p = .585$ (s. Tabelle 22).

Tabelle 22

Vergleich der Aggressionsbereitschaft (Mittelwerte, Standardabweichungen) im FAF (FAF-Agg T-Wert), dem Foto-Hand-Test (FHT-AOS) und dem PFT (PFT-E) zwischen Probanden mit unauffälligen ($> \text{Cut-off}$) und auffälligen ($< \text{Cut-off}$) SN-S-Werten

	SN-S-Gesamtwert unauffällig	SN-S-Gesamtwert auffällig		
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t</i>	<i>p</i>
FAF-Agg	51.6 (9.0)	38.1 (7.1)	3.59	.002
FHT-AOS	0.4 (4.2)	-2.2 (4.3)	1.76	.088
PFT-E	2.7 (2.0)	1.8 (1.9)	1.71	.093
	Skala Aggression unauffällig	Skala Aggression auffällig		
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t</i>	<i>p</i>
FAF-Agg	49.0 (10.7)	38.4 (4.8)	2.13	.047
FHT-AOS	-0.6 (4.5)	-1.0 (4.4)	.232	.818
PFT-E	2.7 (2.1)	1.6 (1.7)	1.86	.069
	Minimierung von Aggression unauffällig	Minimierung von Aggression auffällig		
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t</i>	<i>p</i>
FAF-Agg	54.0 (5.6)	39.6 (9.3)	4.23	.000
FHT-AOS	-0.3 (3.4)	-1.1 (5.0)	.552	.585
PFT-E	3.1 (2.2)	1.8 (1.8)	2.38	.021

In Tabelle 23 sind die Ergebnisse der kategorialen Auswertung von SN-S, FAF-Agg, FHT-AOS und PFT-E dargestellt. Zwar erzielte die Mehrheit der Probanden mit supernormalem Antwortverhalten unterdurchschnittliche Werte im FAF-Agg und im PFT-E und keiner von ihnen überdurchschnittliche Werte, der Zusammenhang zwischen SN-S-Gesamtwert und FAF-Agg ($p = .107$) bzw. PFT-E ($p = .753$) war jedoch jeweils nicht signifikant. Der FHT-

AOS war in der Gruppe der Probanden mit auffälligem SN-S-Gesamtwert tendenziell häufiger unterdurchschnittlich und tendenziell seltener überdurchschnittlich als bei Probanden mit unauffälligem SN-S-Gesamtwert, doch die Mehrheit erzielte durchschnittliche Werte, sodass sich kein signifikanter Zusammenhang zwischen Antwortverhalten und FHT-AOS ergab ($p = .260$).

Tabelle 23

Vergleich der kategorialen Auswertung von FAF-Agg, FHT-AOS und PFT-E zwischen Probanden mit unauffälligen bzw. auffälligen SN-S-Ergebnissen (Fortsetzung S. 88)

SN-S	FAF-Agg (N = 21)			Σ
	unterdurchschnittlich	durchschnittlich	überdurchschnittlich	
GW unauffällig (%)	2 (15.4)	8 (61.5)	3 (23.1)	13 (100)
GW auffällig (%)	5 (62.5)	3 (37.5)	0	8 (100)
Agg unauffällig (%)	4 (25.0)	9 (56.2)	3 (18.8)	16 (100)
Agg auffällig (%)	3 (60.0)	2 (40.0)	0	5 (100)
MA unauffällig (%)	0	8 (80.0)	2 (20.0)	10 (100)
MA auffällig (%)	7 (63.6)	3 (27.3)	1 (9.1)	11 (100)

	FHT-AOS (N = 35)			Σ
	unterdurchschnittlich	durchschnittlich	überdurchschnittlich	
GW unauffällig (%)	2 (10.0)	10 (50.0)	8 (40.0)	20 (100)
GW auffällig (%)	4 (26.7)	9 (60.0)	2 (13.3)	15 (100)
Agg unauffällig (%)	4 (16.7)	13 (54.2)	7 (29.2)	24 (100)
Agg auffällig (%)	2 (18.2)	6 (54.5)	3 (27.3)	11 (100)
MA unauffällig (%)	1 (6.7)	9 (60.0)	5 (33.3)	15 (100)
MA auffällig (%)	5 (25.0)	10 (50.0)	5 (25.0)	20 (100)

Fortsetzung
Tabelle 23

	PFT-E (N = 55)			Σ
	unterdurchschnittlich	durchschnittlich	überdurchschnittlich	

	schnittlich	schnittlich	schnittlich	
GW unauffällig (%)	20 (64.5)	10 (32.3)	1 (3.2)	31 (100)
GW auffällig (%)	18 (75.0)	6 (25.0)	0	24 (100)
Agg unauffällig (%)	23 (63.9)	12 (33.3)	1 (2.8)	36 (100)
Agg auffällig (%)	15 (78.9)	4 (21.1)	0	19 (100)
MA unauffällig (%)	12 (57.1)	8 (38.1)	1 (4.8)	21 (100)
MA auffällig (%)	26 (76.5)	8 (23.5)	0	34 (100)

Ein ähnliches Muster zeigte sich für die Subskala Aggression. Probanden mit auffälligen Werten erzielten häufiger unterdurchschnittliche und nie überdurchschnittliche FAF-Agg- und PFT-E-Werte, der Zusammenhang zwischen dieser Skala und FAF-Agg ($p = .376$) bzw. PFT-E ($p = .585$) war jeweils aber nicht signifikant. Im FHT-AOS unterschieden sich die Probanden mit unauffälligem und auffälligem Antwortverhalten ebenfalls nicht signifikant ($p = 1.0$).

Die Probanden, die auf der Skala Minimierung von Aggression einen auffälligen Wert erzielten, hatten signifikant häufiger unterdurchschnittliche FAF-Agg-Werte als Probanden mit unauffälligen Werten ($p < .01$), die alle durchschnittliche oder überdurchschnittliche FAF-Agg-Werte erzielten. Auch im PFT-E wurden tendenziell häufiger unterdurchschnittliche Werte erreicht, allerdings war der Zusammenhang zwischen Antwortverhalten und PFT-E nicht signifikant ($p = .167$). Der FHT-AOS fiel für Probanden mit Minimierung von Aggression zwar ebenfalls häufiger unterdurchschnittlich und seltener überdurchschnittlich aus, aber auch hier war der Zusammenhang nicht signifikant ($p = .429$).

Von den 16 Probanden, von denen FAF und FHT vorlagen (vgl. Tabelle 24), erzielten sechs (38%) übereinstimmend einen durchschnittlichen Wert, aber keiner in beiden Verfahren übereinstimmend einen über- oder unterdurchschnittlichen Wert. Von drei Probanden mit durchschnittlichem FAF-Agg-Wert erzielten zwei einen überdurchschnittlichen, einer einen unterdurchschnittlichen FHT-AOS. Ein Proband mit überdurchschnittlichem FAF-Agg-Wert hatte einen durchschnittlichen FAF-AOS. Es zeigte sich kein signifikanter Zusammenhang zwischen dem SN-S-Gesamtwert und der Übereinstimmung von FAF-Agg und FHT-AOS ($p = .119$). Zwar dissimulierten fünf von sechs Probanden mit auffälligem SN-S-Gesamtwert ihre Aggressionsbereitschaft im FAF, aber auch drei der zehn Probanden mit unauffälligem SN-S-Gesamtwert. Auch mit der Skala Aggression zeigte sich kein signifikanter Zusammenhang ($p = .108$), obwohl alle vier Probanden mit auffälligem Skalenwert ihre Aggressionsbereitschaft dissimulierten. Allerdings zeigten auch vier Proban-

den mit unauffälligem Skalenwert eine Diskrepanz zwischen FAF-Agg und FHT-AOS i.S. einer Dissimulation. Dagegen wiesen Probanden, die auf der Skala Minimierung von Aggression auffällige Werte erzielten, signifikant häufiger eine Diskrepanz zwischen FAF-Agg und FHT-AOS auf als Probanden mit unauffälligen Werten auf dieser Skala ($p < .01$). Keiner der neun Probanden mit auffälligem Antwortverhalten zeigte eine Aggravation bezüglich der Aggressionsbereitschaft, aber acht dissimulierten diesbezüglich. Keiner von den sieben Probanden mit unauffälligem Antwortverhalten dissimulierte seine Aggressionsbereitschaft, bei sechs stimmten FAF-Agg und FHT-AOS überein, einer aggravierte.

Tabelle 24

Diskrepanz zwischen selbstberichteter und mithilfe (semi-)projektiver Verfahren erfasster Aggressionsbereitschaft (Aggravation, Übereinstimmung, Dissimulation) in Abhängigkeit von den SN-S-Ergebnissen (Fortsetzung S. 90)

	Übereinstimmung FAF und PFT ($N = 17$)			Σ
	Aggravation	Übereinstimmung	Dissimulation	
GW unauffällig (%)	7 (70.0)	3 (30.0)	0	10 (100)
GW auffällig (%)	4 (57.1)	3 (42.9)	0	7 (100)
Agg unauffällig (%)	9 (69.2)	4 (30.8)	0	13 (100)
Agg auffällig (%)	2 (50.0)	2 (50.0)	0	4 (100)
MA unauffällig (%)	6 (75.0)	2 (25.0)	0	8 (100)
MA auffällig (%)	5 (55.6)	4 (44.4)	0	9 (100)

Fortsetzung Tabelle 24	Übereinstimmung FAF und FHT ($N = 16$)			
	Aggravation	Übereinstimmung	Dissimulation	Σ
GW unauffällig (%)	1 (10.0)	6 (60.0)	3 (30.0)	10 (100)
GW auffällig (%)	0	1 (16.7)	5 (71.4)	6 (100)
Agg unauffällig (%)	1 (8.3)	7 (58.3)	4 (33.3)	12 (100)
Agg auffällig (%)	0	0	4 (100)	4 (100)
MA unauffällig (%)	1 (14.3)	6 (85.7)	0	7 (100)
MA auffällig (%)	0	1 (11.1)	8 (88.9)	9 (100)

Es gab keinen Probanden, von dem ein FAF vorlag und der im PFT eine überdurchschnittliche Anzahl von E-Antworten gegeben hatte. Vielmehr erzielten 12 der 17 Probanden, von denen beide Instrumente vorlagen, einen unterdurchschnittlichen PFT-E-Wert. Die Hälfte von ihnen hatte im FAF einen unterdurchschnittlichen AGG-Wert, die andere Hälfte einen durchschnittlichen. Von den fünf Probanden mit durchschnittlichem PFT-E-Wert hatten vier auch einen durchschnittlichen, einer einen überdurchschnittlichen FAF-Agg-Wert. Dementsprechend zeigte sich kein Zusammenhang zwischen dem Antwortverhalten und der Diskrepanz hinsichtlich selbstberichteter und im semiprojektiven PFT gezeigter Aggressionsbereitschaft (SN-S-Gesamtwert: $p = .644$; Skala Aggression: $p = .584$; Skala Minimierung von Aggression: $p = .620$; vgl. Tabelle 24).

5.7.2 Persönlichkeitsstörungen

5.7.2.1 Anzahl unterdurchschnittlicher Skalenwerte in klinischen Persönlichkeitsinventaren

Der SN-S-Gesamtwert korrelierte hoch negativ und signifikant mit der Anzahl unterdurchschnittlicher Skalen im IKP-Eg, $r(25) = -.82$, $p < .001$, sowie moderat negativ und signifikant mit der Anzahl unterdurchschnittlicher Skalen im IKP-G, $r(27) = -.44$, $p < .05$. Mit der Anzahl unterdurchschnittlicher Skalen im PSSI zeigte sich ebenfalls ein signifikanter moderater negativer Zusammenhang, $r(24) = -.43$, $p < .05$. Die Streudiagramme können dem Anhang entnommen werden (Anhang C, Abbildung 9).

Die Probanden mit supernormalem Antwortverhalten ($n = 11$) erzielten eine signifikant höhere Anzahl an unterdurchschnittlichen Skalen im IKP-G ($M = 2.6$, $SD = 2.2$) als Probanden mit unauffälligem Antwortverhalten ($n = 16$, $M = 0.9$, $SD = 1.5$), $t(25) = -2.32$, $p < .05$.

.05. Auch im IKP-Eg erzielten supernormal antwortende Probanden ($n = 9$) signifikant mehr unterdurchschnittliche Skalenwerte ($M = 5.3$, $SD = 2.2$) als offen antwortende Probanden ($n = 15$, $M = 0.4$, $SD = 1.1$), $t(10) = -6.17$, $p < .001$. Im PSSI erzielten Probanden mit auffälligen SN-S-Gesamtwerten ($n = 11$) zwar ebenfalls tendenziell mehr unterdurchschnittliche Skalenwerte ($M = 3.6$, $SD = 2.4$) als Probanden mit unauffälligen SN-S-Gesamtwerten ($n = 13$, $M = 2.1$, $SD = 2.1$), der Unterschied war aber nicht signifikant, $t(22) = -1.72$, $p = .099$.

5.7.2.2 Diskrepanz zwischen klinischer Einschätzung und Selbstbericht

Bei 29 Probanden war vom psychiatrischen Sachverständigen eine Persönlichkeitsstörung diagnostiziert worden. Von ihnen zeigten 16 (55%) supernormales Antwortverhalten, 13 (45%) erzielten in der SN-S einen unauffälligen Wert. Sie unterschieden sich damit nicht signifikant von den Probanden ohne Persönlichkeitsstörung, $\chi^2(1, N = 108) = 2.19$, $p = .139$. Denn von diesen erzielten 48 (61%) einen unauffälligen und 31 (39%) einen auffälligen SN-S-Gesamtwert.

Auch hinsichtlich des Gruppenmittelwerts fanden sich keine signifikanten Unterschiede, $t(106) = 1.30$, $p = .197$. Die Probanden mit Persönlichkeitsstörung erzielten in der SN-S im Mittel 62.9 Punkte ($SD = 19.27$), die Probanden ohne Persönlichkeitsstörung 68.3 Punkte ($SD = 18.9$).

Von 21 Probanden mit der Diagnose Persönlichkeitsstörung lagen Ergebnisse aus klinischen Persönlichkeitsinventaren vor. Von ihnen erzielten acht (38%) auf der mit der Diagnose übereinstimmenden Skala einen überdurchschnittlichen Wert. Alle hatten in der SN-S unauffällige Werte. Bei den anderen 13 Probanden (62%) stimmte die klinische Diagnose nicht mit dem Selbstbericht überein. Darunter waren vier Probanden (31%) mit unauffälligen und neun Probanden (69%) mit auffälligen SN-S-Gesamtwerten. Bei Probanden mit supernormalem Antwortverhalten stimmten Selbsteinschätzung und klinische Diagnose signifikant häufiger nicht überein ($p < .01$).

6 Diskussion

Ziel der vorliegenden Arbeit war eine Überprüfung der Reliabilität und Validität der überarbeiteten deutschen Fassung der SN-S, außerdem die Untersuchung der Häufigkeit supernormalen Antwortverhaltens bei forensischen Begutachtungen sowie dessen Auswirkungen auf die Befunde testpsychologischer Verfahren. Dazu wurden die Daten einer Kontrollstichprobe und die Daten von Probanden, die in der Abteilung für Forensische Psychiatrie der LMU begutachtet wurden, analysiert. Die Diskussion des methodischen Vorgehens und der sich daraus ergebenden Grenzen der Studie bzw. der Schlussfolgerungen für die Praxis erfolgt nach Zusammenfassung und Diskussion der einzelnen Ergebnisse.

6.1 Zusammenfassung und Diskussion der Ergebnisse

6.1.1 Reliabilitätsanalyse und Bildung von Subskalen

Die interne Konsistenz der SN-S war mit $\alpha = .92$ hoch und besser als von Cima et al. (2008) und von Veress (2010) berichtet. In den Substichproben der Schuldfähigkeits- und Zahlungsbegehrenprobanden, die höhere Streuungen der Testkennwerte aufwiesen, war die interne Konsistenz etwas höher ($\alpha = .95$) als in den anderen Substichproben mit geringerer Streuung der Werte (KG: $\alpha = .86$, Fake-KG: $\alpha = .73$), lag jedoch dort auch in einem guten bzw. akzeptablen Bereich. Wie auch von Veress (2010) berichtet, waren die mittleren Inter-Item-Korrelationen (MIC) sowohl in der Gesamtstichprobe als auch in den Substichproben schwach, was zu einem gewissen Grad auf Varianzeinschränkungen zurückzuführen sein dürfte, denn die MIC fielen in den Stichproben mit größerer Streuung höher aus, zum anderen aber vor allem darauf, dass die SN-S keine homogene Skala ist. Es werden sowohl soziale Erwünschtheit als auch Beschwerdenminimierung bzw. Dissimulation (positive Antwortverzerrungen) erfasst; aber auch negative Antwortverzerrungen beeinflussen die Testergebnisse (s. Kap. 6.1.4). Zudem ist davon auszugehen, dass die Ergebnisse auch von der tatsächlich vorhandenen Beschwerdeblastung beeinflusst werden, wie es von Skalen für negative Antwortverzerrungen bekannt ist (Mader, 2014; Wetzler & Marlowe, 1990).

Dementsprechend waren die MIC der Subskalen, bei denen eine größere Homogenität anzunehmen ist als bei dem Gesamtwert, etwas höher, mit Ausnahme der Skala Affektive Symptome jedoch überwiegend schwach ausgeprägt ($\text{MIC} < .40$). Die Reliabilität der Subskalen Affektive Symptome ($\alpha = .84$), Psychotische Symptome ($\alpha = .73$) und Aggression

($\alpha = .78$) war als akzeptabel bis gut zu beurteilen, die übrigen Skalen wiesen unzureichende Werte auf ($\alpha < .70$), was angesichts der geringeren Itemzahl, aus denen sich die Skalen zusammensetzen (4-5 Items), nicht überrascht.

Mit der geringen Homogenität der Subskalen übereinstimmend zeigte die Faktorenanalyse, dass eine sinnvolle Zuordnung der 34 Items zu sieben Subskalen nicht möglich ist. Lediglich drei der extrahierten Faktoren ließen sich sinnvoll interpretieren und konnten am ehesten als Minimierung depressiver Beschwerden, Minimierung von Aggression und Minimierung von ungewöhnlichem (dissoziativem und psychotischem) Erleben beschrieben werden. Diese drei Komponenten setzen sich aus mehr Items zusammen als die ursprünglichen sieben Subskalen. Die Reliabilität bzw. interne Konsistenz der Skalen MDe (9 Items), MA (8 Items) und MUE (6 Items) fiel dementsprechend höher aus ($\alpha = .76-.88$).

6.1.2 Cut-off-Werte und diagnostische Validität

Trotz des systematischen Stichprobenausfalls in den Kontrollstichproben (s. Kap. 6.2) lag der optimale Cut-off der SN-S bei 61 Punkten, was mit dem von Veress (2010) errechneten Wert und dem optimalen Cut-off der Originalversion (Cima et al., 2008) übereinstimmte. Zwar ist davon auszugehen, dass die SN-S eine valide Einstufung ermöglicht (AUC = .86). Sensitivität und negativer Vorhersagewert waren mit 94% bzw. 96% höher als von Cima et al. (2008) berichtet, Spezifität und positiver Vorhersagewert mit 69% bzw. 62% aber nicht zufriedenstellend. Hier ist jedoch zu berücksichtigen, dass der hohe Anteil falsch positiver Treffer auch auf das Vorgehen bei der Stichprobenkorrektur (s. Kap. 6.2) zurückzuführen ist und die errechneten Koeffizienten somit eine Mindestschätzung der Spezifität bzw. des positiven Vorhersagewerts darstellen. Ausgehend von der fehlenden Eindimensionalität der Skala ist es aber auch möglich, dass viele Kontrollprobanden aufgrund einer geringen Beschwerdebelastung niedrigere Werte erzielten. Dafür spricht, dass in der KG am häufigsten – von mehr als der Hälfte (57%) – Werte unterhalb des Cut-offs erzielt wurden. Inwieweit die SN-S-Ergebnisse durch das Ausmaß psychischer Beschwerden beeinflusst werden, sollte daher in künftigen Validierungsstudien untersucht werden. Für forensische Begutachtungen bedeuten die Ergebnisse, dass bei unauffälligem Abschneiden in der SN-S das Vorliegen von Dissimulationstendenzen mit relativ hoher Sicherheit ausgeschlossen werden kann, ein positiver SN-S-Wert jedoch nicht zwangsläufig auf übermäßige Antwortverzerrungen schließen lässt.

Aufgrund der insgesamt schlechten Reliabilitätswerte waren schlechtere Validitätsindizes für die SN-S-Subskalen zu erwarten. Zwar waren die AUC-Werte für alle Skalen ausrei-

chend bis gut ($AUC = .72-.82$), die Gesamttrefferquoten jedoch nicht zufriedenstellend (59%-78%). Ähnlich wie beim SN-S-Gesamtwert waren bei vier Subskalen (Soziale Erwünschtheit, Affektive Symptome, Dissoziatives Erleben, Psychotische Symptome) Sensitivität und negativer Vorhersagewert deutlich größer als Spezifität und positiver Vorhersagewert. Die Skalen Zwang und Aggression wiesen dagegen jeweils eine höhere Spezifität sowie eine höhere Rate falsch negativer Klassifikationen (46% bzw. 34%) auf. Möglicherweise ist dies darauf zurückzuführen, dass die Probanden der Fake-KG Zwanghaftigkeit als tendenziell eher erwünscht beurteilten und daher solche Merkmale seltener verneinten. Aggressivität sahen sie möglicherweise als typisches Merkmal für Straftäter an und verneinten die entsprechenden Items daher seltener.

Die aus der Faktorenanalyse extrahierten Skalen wiesen ebenfalls unzureichende Werte auf. Zwar waren die AUC-Werte etwas höher als die der korrespondierenden Subskalen (Affektive Symptome, Aggression, Ungewöhnliches Erleben), zufriedenstellend fiel jedoch jeweils lediglich der negative Vorhersagewert aus (81%-95%). In der Stichprobe zur Kreuzvalidierung waren die Indizes noch niedriger, da sowohl die Rate falsch Positiver (21%-43%) als auch die Rate falsch Negativer (21%-53%) jeweils beträchtlich waren. Eine quantitative Interpretation der SN-S Subskalen ist daher bei Begutachtungen nicht zu empfehlen. Die Werte können allenfalls für eine qualitative Auswertung i.S. einer Diskrepanzdiagnostik genutzt werden, um Dissimulationstendenzen auszuschließen, etwa wenn ein Gewaltstraftäter auf den Skalen Aggression oder Minimierung von Aggression jeweils Werte oberhalb des Cut-offs erzielt.

6.1.3 Konvergente Validität

Mit der PPI-Skala Unaufrichtige Beantwortung zeigte sich lediglich eine geringe dimensionale ($r = -.16$) und kategoriale Übereinstimmung (40%). Die Stärke des Zusammenhangs war deutlich geringer als die von Eisenbarth und Alpers (2007) berichtete Korrelation zwischen der Skala Unaufrichtige Beantwortung und einer Skala zu sozialer Erwünschtheit. Mit der PPI-Skala Soziale Erwünschtheit fielen die Zusammenhänge dagegen erwartungsgemäß etwas höher, jedoch ebenfalls schwach aus ($r = -.36$). Im Vergleich zur Klassifikationsübereinstimmung mit der UA-Skala war der Anteil an Probanden, die in der SN-S auffällige, auf der PPI-Skala aber unauffällige Werte erzielten, deutlich geringer. Der Anteil mit unauffälligem SN-S-Gesamtwert, jedoch auffälligem Wert auf der PPI-Skala war aber deutlich größer, sodass lediglich bei 62% die Klassifikation des Antwortverhaltens übereinstimmte. Dabei ist allerdings zu berücksichtigen, dass die statistischen Zusammen-

hänge dadurch gemindert sind, dass bei der Stichprobenkorrektur aus KG Probanden entfernt wurden, die übereinstimmend auffällige Werte erzielt hatten und aus Fake-KG Probanden, die übereinstimmend unauffällige Werte erzielt hatten. Sehr hohe Zusammenhänge waren nicht erwartet worden, da die SN-S im Gegensatz zu den PPI-Kontrollskalen neben sozialer Erwünschtheit vor allem Dissimulationstendenzen erfasst. Damit übereinstimmend zeigten beide PPI-Skalen jeweils etwas stärkere Zusammenhänge mit den SN-S-Subskalen Soziale Erwünschtheit, Aggression (AGG) und Minimierung von Aggression (MA) und die schwächsten Zusammenhänge mit den Skalen Dissoziatives Erleben, Psychotische Symptome und Minimierung von ungewöhnlichem Erleben.

Aufgrund der geringen Stichprobengrößen waren keine signifikanten Zusammenhänge zwischen den SN-S-Skalen und den anderen erfassten Kontrollskalen erwartet worden. Die Ergebnisse haben lediglich explorativen Charakter. Die stärksten dimensionalsten Übereinstimmungen i.S. von überwiegend moderaten bis hohen Korrelationen zeigten die SN-S-Skalen mit den Kontrollskalen des MMPI, die ebenfalls Dissimulationstendenzen erfassen und bei der Konstruktion der ursprünglichen Fassung der SN-S herangezogen wurden. Die Korrelationen fielen deutlich höher aus als für die 44-Item Vorgängerversion von Cima et al. (2003a). Mit den Kontrollskalen, die vorwiegend soziale Erwünschtheit erfassen, zeigten sich überwiegend schwache bis moderate Zusammenhänge. Dies entspricht dem Ergebnis von Veress (2010), die ebenfalls lediglich einen zwar signifikanten, aber schwachen Zusammenhang mit der Offenheitsskala des Freiburger Persönlichkeitsinventars (Fahrenberg et al., 2010) fand.

Dass gerade die Offenheitsskala des FAF, die durchgehend (sehr) schwache Zusammenhänge mit den SN-S-Skalen hatte, am geringsten mit den Subskalen zur Minimierung von Aggression korrelierte, lässt ihre Eignung als Kontrollskala für ein Selbstauskunftsverfahren zur Aggressionsbereitschaft fraglich erscheinen, wie es bereits von Biedermann (2007) berichtet wurde.

Die Analysen zur kategorialen Übereinstimmung zeigten, dass es kaum Probanden gab, die in der SN-S unauffällige, aber auf anderen Kontrollskalen auffällige Werte erzielten, was in Übereinstimmung mit der hohen Sensitivität des Instruments steht, die höher ist als etwa die der Kontrollskalen des MMPI, die z.B. bei Baer, Wetter, Nichols, Greene & Berry (1995) zwischen 78 und 83% lag.

Aus diesen Ergebnissen lässt sich für die Praxis die Schlussfolgerung ableiten, dass eine Skala alleine nicht ausreicht, um das Antwortverhalten eines Probanden zuverlässig zu klassifizieren. So hatte z.B. ein Proband in der SN-S einen auffälligen Wert und erzielte

auf der MMPI-L-Skala einen Wert im Bereich einer möglichen Simulation. Möglicherweise lag hier ein hybrider Antwortstil (Rogers, 2008a) vor. Solche Ergebnisse unterstreichen die Bedeutsamkeit einer multimodalen Diagnostik von Antwortverzerrungen.

Insgesamt liefern die Befunde Hinweise dafür, dass die SN-S sowohl soziale Erwünschtheit als auch Dissimulationstendenzen erfasst, die Ergebnisse sollten allerdings an einer größeren Stichprobe repliziert werden.

6.1.4 Erfassung von negativen Antwortverzerrungen mit der SN-S

Es kann angenommen werden, dass Probanden mit negativen Antwortverzerrungen in der SN-S mehr bzw. eine stärkere Ausprägung der erfassten Beschwerden berichten, sodass sie besonders hohe Gesamtwerte erzielen. Dann muss die dichotome Auswertung wie bei den MMPI-Kontrollskalen durch eine dreistufige (unauffällig vs. dissimulierend vs. aggravierend) ersetzt werden. Der SN-S-Gesamtwert korrelierte nicht nur signifikant stark mit den MMPI-Kontrollskalen (s.o.), sondern auch mit dem Gesamtwert des SFSS ($r = .78$). Ebenso fanden sich signifikante moderate bis hohe Zusammenhänge des SFSS mit den Subskalen der SN-S ($r = .41-.81$). Während die Korrelationen mit den SFSS Subskalen Affektive Symptome, Gedächtnisstörungen und Neurologische Beeinträchtigungen durchweg signifikant moderat bis hoch ausfielen, zeigte die Skala Psychotische Symptome lediglich mit der SN-S-Skala Psychotische Symptome einen signifikanten hohen Zusammenhang, mit den anderen Skalen durchweg signifikante moderate Zusammenhänge. Die Skala Niedrige Intelligenz wies lediglich schwache oder gar keine Zusammenhänge mit den SN-S-Skalen auf. Dies ist zum einen damit zu begründen, dass Auffälligkeiten in diesem Bereich mit der SN-S nicht erfasst werden. Zum anderen handelt es sich bei den SFSS-Skalen Psychotische Symptome und Niedrige Intelligenz um die Skalen, auf denen von Begutachtungsprobanden und stationär psychiatrischen Patienten seltener hohe Werte erzielt werden (Mader, 2014), was zu Varianzeinschränkungen und damit Minderung der Korrelationen führt.

Mit dem SFSS-Gesamtwert als Außenkriterium zeigte die SN-S zwar eine hohe Vorhersagegüte ($AUC = .91$), bei einem Cut-off-Wert von 78 Punkten waren Spezifität (90%) und negativer Vorhersagewert (93%) gut, die Sensitivität mit 78% und der positive Vorhersagewert mit 69% aber unzureichend. Dabei ist jedoch zu berücksichtigen, dass der SFSS selbst in der Studie von Cima et al. (2003b) lediglich eine moderate diagnostische Validität mit Indizes in Höhe von 86% (Spezifität, negativer Vorhersagewert) bzw. 87% (Sensitivität, positiver Vorhersagewert) hatte. Es bedarf daher weiterer Studien mit anderen Außen-

kriterien, um die diagnostische Validität der SN-S bei der Erfassung negativer Antwortverzerrungen zu prüfen. Dennoch kann bereits jetzt bei Begutachtungen i.S. einer multimodalen Diagnostik der SN-S-Gesamtwert herangezogen werden, um negative Antwortverzerrungen auszuschließen, wenn sowohl SFSS- als auch SN-S-Gesamtwert unauffällig sind.

6.1.5 Auftretenshäufigkeit von Supernormalität in Abhängigkeit vom Untersuchungskontext

Die Hypothese, dass die Häufigkeit supernormalen Antwortens von der Fragestellung der Begutachtung abhängt, konnte nur bedingt bestätigt werden. Sowohl für den SN-S-Gesamtwert als auch für die Subskalen zeigten sich jeweils signifikante Häufigkeits- und Mittelwertsunterschiede. Rein deskriptiv war die Rangfolge der Untersuchungsgruppen konsistent mit der Hypothese: Zahlungsbegehrenprobanden erzielten höhere SN-S-Gesamtwerte ($M = 77.6$) und antworteten seltener supernormal (22%) als Schuldfähigkeitsprobanden ($M = 71.5$; 38% supernormal), gefolgt von Prognoseprobanden ($M = 60.1$; 51% supernormal) und Gesundheitsnachweisprobanden ($M = 52.1$; 88% supernormal). Die niedrigsten Werte wurden von der Fake-KG erzielt ($M = 51.8$; 94% supernormal), die ähnliche Werte erzielte wie die Gesundheitsnachweisprobanden. Die Probanden der KG ($M = 66.8$; 31% supernormal) erzielten Werte, die etwas niedriger als die der Schuldfähigkeits-, aber höher als die der Prognoseprobanden waren. Signifikante hypothesenkonforme Unterschiede zeigten sich lediglich bei der dimensionalen Auswertung, wobei Zahlungsbegehren- und Schuldfähigkeitsprobanden, die sich nicht signifikant voneinander unterschieden, signifikant höhere Werte erzielten als Prognose- und Gesundheitsnachweisprobanden. Bei der kategorialen Auswertung zeigten sich signifikante hypothesenkonforme Unterschiede lediglich zwischen Gesundheitsnachweis- und Zahlungsbegehrenprobanden.

Für die Subskalen Affektive Symptome, Zwang, Psychotische Symptome, Dissoziatives Erleben, Minimierung von depressiven Beschwerden und Minimierung von Aggression zeigte sich ein ähnliches Muster, wobei sich bei der kategorialen Auswertung die Zahlungsbegehrenprobanden auch z.T. signifikant von den Prognoseprobanden unterschieden. Auf der Skala Soziale Erwünschtheit unterschieden sich sowohl bei der kategorialen als auch bei der dimensionalen Auswertung dagegen lediglich die Gesundheitsnachweisprobanden von den anderen Begutachtungsprobanden. Die Skala Minimierung von ungewöhnlichem Erleben war die einzige, bei der sich die Zahlungsbegehrenprobanden hinsichtlich des Mittelwerts kaum von den anderen Begutachtungsprobanden unterschieden und sogar häufiger auffällige Werte erzielten als Schuldfähigkeitsprobanden, was darauf zurückzu-

führen war, dass psychotische und dissoziative Symptome von diesen Probanden am häufigsten verneint wurden (48% bzw. 35%).

Dass sich bei der dimensionalen Auswertung, nicht aber bei der kategorialen Auswertung signifikante Unterschiede zwischen Schuldfähigkeits- und Prognoseprobanden zeigen, ist konsistent mit den Ergebnissen von Zimmermann (2012). Allerdings ist in beiden Studien der Stichprobenumfang der Untersuchungsgruppen gering, was die fehlende Signifikanz trotz deutlicher deskriptiver Unterschiede erklären kann. Insgesamt sprechen die Befunde dafür, dass weniger die Fragestellung der Begutachtung per se, sondern das individuelle Ziel, das der Proband bei der Begutachtung verfolgt (Greene, 2007), entscheidend für das Auftreten von Antwortverzerrungen ist. Denn bei den beiden Gruppen mit den deutlichsten Unterschieden, Zahlungsbegehren und Gesundheitsnachweis, handelt es sich um diesbezüglich homogenere Stichproben als bei den Schuldfähigkeitsprobanden. Letztere können nicht nur bestrebt sein, wegen einer psychischen Störung Strafmilderung zu erhalten, sondern auch, sich möglichst unauffällig und gesund zu präsentieren, insbesondere wenn sie die ihnen zur Last gelegte Tat leugnen oder eine Unterbringung im psychiatrischen Maßregelvollzug vermeiden wollen. Die geringere Homogenität dieser Subgruppe zeigte sich auch in einer größeren Streuung der Testwerte. Für weitere Untersuchungen sollten also die Subgruppen eher nach dem Kriterium des individuellen Ziels als nach der Fragestellung der Begutachtung gebildet werden. Dafür spricht auch, dass die Prognoseprobanden, die i.d.R. bestrebt sind, sich gesund und unauffällig zu präsentieren, um aus der Haft oder dem Maßregelvollzug entlassen zu werden, bezüglich des Antwortverhaltens den Probanden, die einen Gesundheitsnachweis anstrebten, ähnlicher waren als den Schuldfähigkeitsprobanden.

Die Häufigkeit positiver Antwortverzerrungen war in der untersuchten Stichprobe höher als in der Literatur angegeben (20-30%; Baer & Miller, 2002; Grossman & Wasyliv, 1988; Heilbrun et al, 1990). Insgesamt hatten 44% der untersuchten Probanden einen auffälligen SN-S-Wert erzielt. Dies dürfte zum einen auf die unterschiedliche Zusammensetzung der untersuchten Stichproben zurückzuführen sein, zum anderen aber auch auf die hohe Sensitivität und geringe Spezifität der SN-S.

6.1.6 Auswirkung von Supernormalität auf testpsychologische Befunde

In Übereinstimmung mit Untersuchungen von anderen Skalen zu positiven Antwortverzerrungen bzw. Studien mit experimentellen Designs an forensischen Stichproben (Grossman et al., 2002; Fiduccia, 2011; Haywood et al., 1993; McEwan et al., 2009; Steffan et al.,

2007) konnte anhand der hier untersuchten Stichprobe gezeigt werden, dass Supernormalität im Begutachtungskontext zu verzerrten Testergebnissen bei Selbstbeurteilungsverfahren führt. Die Ergebnisse sprechen für die Brauchbarkeit der SN-S bei der Aufdeckung von positiven Antwortverzerrungen.

Die Hypothese, dass Probanden mit supernormalem Antwortverhalten eine geringere Aggressionsbereitschaft berichten, konnte bestätigt werden und zwar sowohl für den SN-S-Gesamtwert als auch die beiden Subskalen Aggression (AGG) und Minimierung von Aggression (MA). Bei der kategorialen Auswertung zeigten sich signifikante Zusammenhänge zwischen dem Antwortverhalten und dem FAF-Ergebnis aber lediglich für die Skala Minimierung von Aggression (MA), was dafür spricht, dass diese mittels der Faktorenanalyse extrahierte Skala die höchste Sensitivität bezüglich der Minimierung von Aggressionsbereitschaft hat (s.u.).

Die Befunde stützen außerdem die Annahme, dass projektive Verfahren weniger leicht zu verfälschen sind als Selbstauskunftsverfahren (Biedermann, 2007; Sewel, 2008; Weiner, 2007), allerdings nur wenn die Messintention nicht zu durchschauen ist. Denn der FHT-AOS-Wert korrelierte nicht mit den SN-S-Skalen und es zeigten sich keine signifikanten Gruppenunterschiede zwischen Probanden mit auffälligen und unauffälligen Werten auf den SN-S-Skalen. Dagegen wies der PFT, bei dem aufgrund der dargestellten Konfliktsituationen von einer größeren Durchschaubarkeit auszugehen ist, hohe Korrelationen mit den SN-S-Werten auf und Probanden mit auffälligen Werten auf der Skala Minimierung von Aggression erzielten häufiger unterdurchschnittliche E-Werte. Dies entspricht früheren Befunden, die zeigen, dass der PFT durchaus anfällig für Verfälschungen ist und dass positive Antwortverzerrungen zu einer Abnahme der E-Antworten führen (King, 1972; Vane, 1954).

Dafür, dass der PFT leichter zu verfälschen ist, sprach auch, dass die Hypothese „Bei Probanden mit supernormalem Antwortverhalten zeigt sich häufiger eine Diskrepanz zwischen projektiven Verfahren und Selbstbericht“ nur für den FHT und die Skala Minimierung von Aggression bestätigt werden konnte. Von neun Probanden mit auffälligem Wert auf dieser Skala dissimulierten, gemessen am FHT-AOS, acht ihre Aggressionsbereitschaft im FAF. Der hohe Anteil von Probanden, die beim Vergleich zwischen FHT und FAF als dissimulierend eingestuft wurden (50%), erklärt auch die zwar nicht signifikante, aber moderate negative Korrelation zwischen Aggressivität im FAF und FHT-AOS. Dagegen erzielten im PFT 71% der Probanden, bei denen auch der FAF-Wert erhoben wurde, unterdurchschnittliche E-Werte, sodass kein Proband als dissimulierend eingestuft wurde. Wünschenswert

wäre ein weiteres Außenkriterium für die Aggressionsbereitschaft, etwa eine Fremdbeurteilung, wie dies von Biedermann (2007) erhoben wurde, gewesen, um die Befunde zu stützen.

Sowohl die dimensionalen als auch die kategorialen Auswertungen deuten darauf hin, dass die Skala Minimierung von Aggression der beste Indikator für Dissimulationstendenzen bezüglich Aggressionsbereitschaft ist. Dies ist auf die höhere Sensitivität dieser mithilfe der Faktorenanalyse extrahierten Subskala zurückzuführen; zwischen 52% und 61% der bezüglich Aggressionsbereitschaft untersuchten Probanden erreichten hier einen auffälligen Wert, im Gegensatz zu 38% bis 44% der Probanden mit auffälligem SN-S-Gesamtwert und 24% bis 35% Probanden mit auffälligem Wert auf der Skala Aggression. Bei Begutachtungen ist allerdings der dementsprechend geringere positive Vorhersagewert der Subskala Minimierung von Aggression zu berücksichtigen. Für eine abschließende Beurteilung, welche Skala in der Praxis die größte Brauchbarkeit besitzt, sind weitere Studien an einem größeren Probandenkollektiv notwendig. Festzuhalten bleibt, dass die Diagnostik von negativen Konstrukten wie Aggressionsbereitschaft bei Begutachtungen mit möglichst nicht oder kaum durchschaubaren Instrumenten durchgeführt werden sollte, wobei in Zukunft weniger projektive Verfahren als vielmehr die objektiven bzw. indirekten Verfahren, wie der IAT (Werner, 2005), angewendet werden dürften.

Auch die Hypothese, dass Probanden mit supernormalem Antwortverhalten häufiger unterdurchschnittliche Skalenwerte in Inventaren zur Erfassung von Persönlichkeitsstilen und -störungen erzielen, konnte zumindest für IKP-G und IKP-Eg bestätigt werden. Es konnten nicht nur die Ergebnisse von Nedopil et al. (2013) von signifikanten moderaten Zusammenhängen zwischen supernormalem Antworten und der Anzahl unterdurchschnittlicher Skalen repliziert werden. Die Probanden mit auffälligen SN-S-Werten erzielten auch signifikant mehr unterdurchschnittliche Skalenwerte im IKP. Während Probanden mit unauffälligem Antwortverhalten jeweils im Mittel keinen oder einen durchschnittlichen Skalenwert aufwiesen, hatten Probanden mit Dissimulationstendenzen im IKP-G, das 11 Skalen umfasst, im Mittel drei und im IKP-Eg, das 13 Skalen umfasst, im Mittel fünf unterdurchschnittliche Skalenwerte.

Für das PSSI konnte die Hypothese dagegen nur bedingt bestätigt werden. Zwar zeigte sich ebenfalls ein signifikanter moderater Zusammenhang zwischen SN-S-Wert und der Anzahl unterdurchschnittlicher Skalenwerte, der Mittelwertsunterschied zwischen Probanden mit supernormalem und unauffälligem Antwortverhalten war aber nicht signifikant. Daraus ist jedoch nicht zu schließen, dass das PSSI weniger anfällig für Antwortverzerrungen ist, als

das IKP. Die fehlende Signifikanz ist neben dem geringen Stichprobenumfang eher darauf zurückzuführen, dass auch bei Probanden mit unauffälligem SN-S-Wert im Mittel zwei der 14 Skalen unterdurchschnittlich ausgeprägt waren, bei den Probanden mit Supernormalität im Durchschnitt knapp vier Skalen. Eine Beurteilung, welches Verfahren sich für forensische Begutachtungen besser eignet, ist erst möglich, wenn einer größeren Stichprobe von Probanden beide Verfahren vorgelegt und die Befunde mit einer Fremdeinschätzung verglichen werden.

Die Hypothese, dass sich bei Probanden mit positiven Antwortverzerrungen häufiger eine Diskrepanz zwischen Selbsteinschätzung und klinischer Diagnose einer Persönlichkeitsstörung zeigt, konnte ebenfalls bestätigt werden. Bei keinem der neun Probanden mit Supernormalität stimmten die Ergebnisse beider Quellen überein, wohl aber bei 67% der 12 Probanden ohne Antwortverzerrungen. Bei den übrigen 33% dürfte die fehlende Übereinstimmung eher auf eine mangelnde Krankheitseinsicht oder Introspektionsfähigkeit bzw. Selbsttäuschung (Paulhus, 2002) der Probanden als auf eine falsch negative Klassifikation durch die SN-S zurückzuführen sein, da die Rate solcher Fehlklassifikationen gering ist (6%).

Die Ergebnisse zeigen aber, dass nicht davon ausgegangen werden kann, dass Personen mit Persönlichkeitsstörung per se wegen mangelnder Krankheitseinsicht weniger Auffälligkeiten oder Beschwerden berichten. Denn in Übereinstimmung mit den Ergebnissen von Zimmermann (2012) zeigte sich kein Zusammenhang zwischen dem Vorliegen einer Persönlichkeitsstörung und dem Auftreten positiver Antwortverzerrungen. Probanden mit der Diagnose Persönlichkeitsstörung erzielten vergleichbar häufig (45%) unauffällige SN-S-Werte wie Probanden ohne Persönlichkeitsstörung (39%). Damit kann der Einsatz von Selbstbeurteilungsverfahren zu Persönlichkeitsakzentuierungen bzw. -störungen durchaus sinnvoll sein, um ökonomisch Informationen über eine entsprechende Pathologie zu erheben, aber nur dann, wenn keine Antwortverzerrungen vorliegen.

Denn sonst führen solche Inventare bei Begutachtungen zu falschen Ergebnissen. Die Ergebnisse dieser Arbeit verdeutlichen zudem, wie in Kapitel 2.2.2.2 und 2.5 ausgeführt, dass bei empirischen Untersuchungen von Zusammenhängen zwischen Antwortverzerrungen und Persönlichkeitsmerkmalen wie Psychopathie mehr auf Fremdbeurteilungsinstrumente als auf Selbstauskunftsverfahren zurückgegriffen werden sollte.

6.2 Diskussion des methodischen Vorgehens

Wesentliches Gütekriterium für ein Instrument zur Aufdeckung von Antwortverzerrungen ist die diagnostische Validität. Da ein Goldstandard bzw. valider Referenzstandard zur Klassifikation des Antwortverhaltens nicht verfügbar ist, wurde auf ein experimentelles Design mit randomisierter Zuweisung der Probanden einer Stichprobe zu einer von zwei Kontrollgruppen mit Standard- oder Fake-good-Instruktion zurückgegriffen. Solche Analogstudien verfügen zwar über eine hohe interne, aber aufgrund der geringen Ähnlichkeit mit der realen Begutachtungssituation über eine geringe externe Validität und liefern daher lediglich Schätzungen der Validität (Rogers, 2008a).

Die Güte der Validitätsschätzung hängt von der Zusammensetzung der Kontrollstichprobe ab und davon, wie gut es durch die Instruktion gelingt, Antwortverzerrungen zu induzieren bzw. die Kontrollprobanden zu motivieren, wie Begutachtungsprobanden im realen Kontext zu antworten (Rogers, 2008d). Im Gegensatz zur Validierungsstudie von Cima et al. (2008) wurde daher zum einen keine studentische Stichprobe rekrutiert, sondern eine bezüglich Alter, Geschlecht und Bildungsstand heterogene Stichprobe zusammengestellt. Diese ist aufgrund des Rekrutierungsweges (per Email über Mitarbeiter der Abteilung für Forensische Psychiatrie) jedoch nicht bevölkerungsrepräsentativ und unterscheidet sich zudem hinsichtlich Alter, Geschlecht, Familienstand, schulischer und beruflicher Bildung signifikant von Gutachtensprobanden, insbesondere jener, die wegen strafrechtlicher Fragestellungen untersucht werden. Allerdings ist, wie in Kap. 2.2.2.2 ausführlich dargestellt, nicht anzunehmen, dass solche Merkmale systematisch mit Antwortverzerrungen zusammenhängen.

Kritisch zu sehen ist auch die geringe Ähnlichkeit mit einer realen Begutachtungssituation, da die Daten der Kontrollprobanden mit einem Onlinefragebogen erhoben wurden. Diese Erhebungsmethode wurde zum einen aus ökonomischen Gründen gewählt, zum anderen aber vor allem um die wahrgenommene Anonymität zu erhöhen und damit die Wahrscheinlichkeit von Antwortverzerrungen in der Kontrollstichprobe mit Standardinstruktion zu verringern. Denn in der Untersuchung von Veress (2010), welche die Daten der Kontrollprobanden in einer Face-to-Face-Untersuchung erhoben hatte, hatten 77% einen auffälligen Gesamtwert erzielt. Daher wurde auch auf eine Speicherung der IP-Adressen verzichtet, was dazu führte, dass eine Analyse der Unit-Nonresponders (Totalausfälle; Baur & Florian, 2009) nicht möglich war. Diese spielt jedoch bei der Beurteilung der Ausschöpf-

fung der hier untersuchten Stichprobe im Gegensatz zur Rate der vorzeitigen Abbrecher eine untergeordnete Rolle (s.u.).

Die Onlineerhebung hatte den Nachteil, dass abgesehen von den erhobenen demographischen Variablen keine Informationen über die Probanden vorlagen, z.B. bezüglich psychischer Erkrankungen, Erfahrungen mit forensischen Begutachtungen oder gar einer strafrechtlichen Vorgeschichte, sodass diese Variablen bei der Interpretation der Befunde nicht berücksichtigt werden konnten.

Wegen der Anonymität der Befragung konnte auch keine Belohnung für die Teilnahme in Aussicht gestellt werden. Deshalb sollte entsprechend der Empfehlungen von Rogers (2008d) die Motivation durch die Formulierung der allgemeinen Instruktionen für die Teilnehmer gestärkt und damit die Anzahl der Drop-outs reduziert werden. In dieser Untersuchung geschah dies, indem die Bedeutung des Forschungsprojekts für die Befunde von Begutachtungen zur Rückfallprognose, welche wiederholt in den Medien teils heftig kritisiert wurden, betont wurde.

Zwar entschlossen sich insgesamt 290 Personen für die Teilnahme am Projekt, durch Drop-outs und Stichprobenkorrektur reduzierte sich der Stichprobenumfang jedoch auf 201 Probanden, was aber für eine Validierung ausreichend ist (Bühner, 2011). Zudem unterschied sich der Stichprobenumfang der beiden Kontrollgruppen zum einen aufgrund der randomisierten Zuweisung (154 vs. 136 Probanden), zum anderen aufgrund der höheren Drop-out-Rate in der Fake-KG (13%) gegenüber der KG (6%) und dem größeren Ausschluss von Probanden in der Fake-KG (41%) gegenüber der KG (8%) bei der Stichprobenkorrektur (s.u.). Die Drop-out-Rate war zwar gering (Musch & Reips, 2000), allerdings systematisch, da nicht nur in der Fake-KG mehr Probanden ihre Teilnahme vorzeitig beendeten, sondern die Mehrheit von ihnen (82%) dies nach Lesen der Fake-Instruktion tat.

Trotz sorgfältiger Formulierung der Fake-Instruktion, schien diese nicht bei allen Probanden in ausreichendem Maße positive Antwortverzerrungen induziert zu haben. Die Kontrollskala des PPI war in den Onlinefragebogen nicht nur zur Überprüfung der konvergenten Validität integriert worden, sondern auch um i.S. einer Datenbereinigung Probanden mit unaufrichtigem Antwortverhalten aus der KG und Probanden ohne Antwortverzerrungen aus der Fake-KG ausschließen zu können. Durch diese Prozedur reduzierte sich der Stichprobenumfang vor allem in der Fake-KG beträchtlich, von 119 auf 70. Doch auch in der KG mussten 8% der Probanden ausgeschlossen werden, was zeigte, dass trotz Betonung der Anonymität positive Antwortverzerrungen auftraten. Das Vorgehen bei der Stichprobenkorrektur ist nicht unproblematisch, da bekannt ist, dass die PPI-Kontrollskala

zu einem nicht unerheblichen Anteil von Fehlklassifikationen des Antwortverhaltens führt (Edens et al., 2001). Zuletzt untersuchten Anderson, Sellbom Wygant und Edens (2013) die diagnostische Validität der PPI-Kontrollskalen Deviant Responding (DR; negative Antwortverzerrungen) und Virtuous Responding (VR; positive Antwortverzerrungen) anhand einer studentischen Stichprobe mit experimentellem Design. Die VR-Skala wies zwar eine hohe Spezifität auf (98%), aber eine geringe Sensitivität (32%), die Gesamttrefferquote betrug 91%. Die diagnostische Validität der deutschen Fassung der PPI-Kontrollskala Unaufrichtige Beantwortung, bei der im Gegensatz zur Originalfassung positive Antwortverzerrungen und zufälliges Antworten integriert sind, wurde bislang nicht untersucht. In der hier analysierten Stichprobe waren die Gütekriterien nicht zufriedenstellend. Die Reliabilität war, wie zu erwarten, gering ($\alpha = .31-.67$), die Spezifität mit 96% zwar hoch, die Sensitivität aber unzureichend, da lediglich 28% der Fake-KG auffällige Werte erzielten, positiver und negativer Vorhersagewert lagen jeweils bei unter 80%. Auch die PPI-SE Skala wies zwar eine angesichts der Eindimensionalität zu erwartende höhere interne Konsistenz ($\alpha = .68-.77$) auf, aber unzureichende Validitätsindizes. Die Sensitivität war mit 71% höher als die der VR-Skala der Originalfassung des PPI-R, Spezifität (70%) und Gesamttrefferquote (70%) jedoch deutlich geringer. Es ist daher anzunehmen, dass bei der Stichprobenkorrektur vor allem in der KG Probanden verblieben waren, die positive Antwortverzerrungen gezeigt hatten. Bei der Berechnung der Validitätsindizes der SN-S könnte das zu einer erhöhten Rate falsch Positiver und damit zu einer Abnahme der Spezifität und des positiven Vorhersagewerts geführt haben. Die PPI-Kontrollskala war ausgewählt worden, weil es sich um die bei psychologischen Begutachtungen in der Abteilung für Forensische Psychiatrie nach der SN-S am häufigsten eingesetzte Kontrollskala handelt und sie nicht nur Faking-good, sondern auch itemunabhängiges Antworten erfasst. Der Anteil von nachträglich aus der Analyse ausgeschlossenen Datensätzen ist lediglich in der Fake-KG etwas höher als in vergleichbaren Studien (17-34%; Übersicht bei Edens et al., 2001), die allerdings die Ergebnisse eines sog. Manipulationschecks zur Datenbereinigung heranzogen. Beim Manipulationscheck wird direkt im Anschluss an die Testbearbeitung geprüft, ob die Probanden die experimentellen Instruktionen verstanden und entsprechend geantwortet haben (Rogers, 2008d). Wäre anstatt des Integrierens der PPI-UA-Skala in den Onlinefragebogen dieses Vorgehen gewählt worden, hätten die Daten der Kontrollstichprobe aber bei der Berechnung der konvergenten Validität nicht berücksichtigt werden können. Der Einsatz einer Kontrollskala und eines Manipulationschecks hätte zu einer Zunahme der Befragungszeit und damit wahrscheinlich bei vielen zu einem Abnehmen der Motivati-

on zur Teilnahme geführt (Bosnjak & Batinic, 2000). Möglicherweise hätte jedoch zusätzlich zur Einschätzung von Experten ein Pretest anhand einer kleinen Kontrollstichprobe mit anschließender Befragung auf Schwächen bei der experimentellen Instruktion hinweisen können. Dies sollte bei weiteren Validierungsstudien der SN-S berücksichtigt werden. Die Datenbereinigungsverfahren wurde auch auf die Daten von Bantel (2011), die zur Kreuzvalidierung der durch die Faktorenanalyse extrahierten Faktoren herangezogen wurde, angewendet. Dies reduzierte die ohnehin kleine Stichprobe, sodass die entsprechenden Ergebnisse mit Vorsicht zu interpretieren sind bzw. nicht generalisiert werden können.

Für die Beantwortung der übrigen Fragestellungen wurden Daten von Begutachtungsprobanden analysiert, die nicht speziell für die vorliegende Arbeit erhoben wurden. Somit kann von einer hohen externen Validität ausgegangen werden. Zwar handelt es sich nicht um eine Vollerhebung aller Begutachtungsprobanden der Abteilung für Forensische Psychiatrie in diesem Erhebungszeitraum, jedoch um eine repräsentative Stichprobe der testpsychologisch untersuchten Probanden. Mit 108 Personen konnte eine im Vergleich zu ähnlichen Studien bzw. für eine forensische Studie recht große Stichprobe in die Analyse einbezogen werden. Die anhand der Begutachtungsfragestellung gebildeten Substichproben sind allerdings recht klein bzw. deutlich unterschiedlich groß ($n = 8$ bis $n = 43$). Den größten Anteil hatten Prognose- gefolgt von Schuldfähigkeitsprobanden, der Anteil von sozialmedizinischen und zivilrechtlichen Probanden war dagegen geringer. Dies ist nicht nur darauf zurückzuführen, dass positive Antwortverzerrungen bei Prognosebegutachtungen obligatorisch erhoben werden, sondern entspricht der Häufigkeitsverteilung von Begutachtungsaufträgen mit entsprechender Fragestellung in der Abteilung für Forensische Psychiatrie. Für die einzelnen Fragestellungen wurden die Ergebnisse anderer testpsychologischer Verfahren erhoben, allerdings lagen für einige Instrumente nur wenige Daten vor ($N < 30$), sodass die Arbeit in diesem Bereich als explorativ anzusehen ist und die Ergebnisse zunächst an einem größeren Kollektiv repliziert werden müssen. Dies betrifft vor allem die Ergebnisse zur konvergenten Validität der SN-S und zu Auswirkungen von Supernormalität auf testpsychologische Befunde. Die Arbeit sollte jedoch auch die Brauchbarkeit der SN-S untersuchen bzw. der Entscheidung dienen, ob sich ihr weiterer Einsatz bei Begutachtungen in der forensischen Psychiatrie lohnt, weshalb hier trotz geringer Stichprobenumfänge alle Ergebnisse referiert wurden.

6.3 Fazit und Empfehlungen für die Praxis

Die vorliegende Arbeit weist zwar einige methodische Schwächen auf, die die Generalisierbarkeit der Ergebnisse zunächst einschränken bzw. eine Replikation einiger Befunde an größeren Probandenkollektiven notwendig machen, bevor abschließend Schlussfolgerungen für die Praxis gezogen werden können. Die Stärke bzw. der klare Vorteil der Arbeit besteht jedoch darin, dass eine für forensische Untersuchungen relativ große Stichprobe von Begutachtungsprobanden eingeschlossen werden konnte, die unter naturalistischen Bedingungen untersucht wurden. Damit ist eine hohe externe Validität gegeben und die Ergebnisse können eher generalisiert werden, als es bei – insbesondere im angloamerikanischen Raum verbreiteten – Studien mit großen Stichproben studentischer Versuchspersonen der Fall ist.

Die Ergebnisse der Arbeit sprechen für eine zufriedenstellende Validität der SN-S bei der Erfassung positiver Antwortverzerrungen von forensischen Begutachtungsprobanden. Offen bleibt, welche Faktoren neben dem individuellen Ziel des Probanden die SN-S-Werte beeinflussen: Vor allem die Auswirkungen von Selbsttäuschung bzw. mangelnder Krankheitseinsicht und dem Ausmaß tatsächlich vorhandener Beschwerden auf das Abschneiden in der SN-S sollten empirisch untersucht werden.

Testpsychologische Untersuchungen der Persönlichkeit sind fester Bestandteil forensischer Begutachtungen, weil sie eine ökonomische Erhebung verschiedener fragestellungsrelevanter Merkmale bzw. des Selbstbildes von Probanden ermöglichen. Dies gilt jedoch nur, wenn die Befunde nicht durch bewusste Täuschungsversuche verzerrt werden. Die Ergebnisse der vorgelegten Arbeit sprechen für die Brauchbarkeit der SN-S, insbesondere des Gesamtwerts, bei der Diagnostik von positiven Antwortverzerrungen. Die Skalenwerte sollten dagegen nur mit Vorsicht bzw. eher qualitativ interpretiert werden. Die Ergebnisse der SN-S sollten außerdem nicht isoliert betrachtet werden. Insbesondere wenn ein Wert unterhalb des Cut-off-Werts erzielt wird, sollte durch den Einsatz weiterer Verfahren sichergestellt werden, dass es sich nicht um ein falsch positives Ergebnis handelt. Die SN-S kann also stets nur Bestandteil einer umfassenderen, multimodalen Diagnostik des Antwortverhaltens sein. Nur so kann im individuellen Fall festgestellt werden, ob und in welchem Ausmaß Antwortverzerrungen vorliegen. Ist davon auszugehen, dass die Befunde durch Täuschungsversuche verzerrt werden, sollte auf den Einsatz von Selbstbeurteilungsverfahren verzichtet und, sofern verfügbar, auf projektive oder indirekte Verfahren zurückgegriffen werden. Zumindest aber müssen die Testresultate angemessen interpretiert und

der Einfluss des Antwortverhaltens diskutiert werden, um zu einer für den Auftraggeber nachvollziehbaren Beurteilung der Befunde zu kommen.

7 Literaturverzeichnis

- Alpers, G. W. & Eisenbarth, H. (2008). *PPI-R. Psychopathic Personality Inventory-Revised. Deutsche Version. Manual*. Göttingen: Hogrefe.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- Anderson, J. L., Sellbom, M., Wygant, D. B. & Edens, J. F. (2013). Examining the Necessity for and Utility of the Psychopathic Personality Inventory – Revised (PPI-R) Validity Scales. *Law and Human Behavior*, 37, 312-320.
- Andresen, B. (2006). *IKP. Inventar Klinischer Persönlichkeitsakzentuierungen. Dimensionale Diagnostik nach DSM-IV und ICD-10. Manual*. Göttingen: Hogrefe.
- Andrews, D. A. & Bonta, J. (2010). *The Psychology of Criminal Conduct*. New Providence: LexisNexis Matthew Bender.
- Baer, R. A. & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment*, 14, 16-26.
- Baer, R. A., Wetter, M. W., Nichols, D. S., Greene, R. & Berry, D. T. R. (1995). Sensitivity of MMPI-2 validity scales to underreporting of symptoms. *Psychological Assessment*, 7, 419-423.
- Bagby, R. M., Rogers, R. & Buis, T. (1994). Detecting malingered and defensive responding on the MMPI-2 in a forensic inpatient sample. *Journal of Personality Assessment*, 62, 191-203.
- Bantel, D. (2011). *Supernormalität bei forensischen Probanden und ihre Zusammenhänge mit Intelligenz und Psychopathie* (Unveröffentlichte Bachelorarbeit). Universität Salzburg.
- Baur, N. & Florian, M. J. (2009). Stichprobenprobleme bei Online-Umfragen. In N. Jakob, H. Schoen & T. Zerback (Hrsg.), *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung* (S. 109-128). Wiesbaden: VS.
- Belschner, W., Lischke, G. & Selg, H. (1971). *Foto-Hand-Test (FHT) zur Erfassung von Aggressivität*. Freiburg: Karl Alber.
- Berardino, S. D., Meloy, J. R., Sherman, M. & Jacobs, D. (2005). Validation of the psychopathic personality inventory on a female inmate sample. *Behavioral Sciences & the Law*, 23, 819-836.
- Bernstein, E. M. & Putnam F. W. (1986). Development, reliability and validity of a dissociation scale. *Journal of Nervous and Mental Disease*, 174, 727-735.

- Biedermann, T. (2007). *Beiträge zur Aggressionsdiagnostik. Multimethodale psychologische Informationsgewinnung bei jungen inhaftierten Gewalttätern im Vergleich mit sozial unauffälligen Jugendlichen*. Hamburg: Dr. Kovač.
- Bohm, E. (2004). *Lehrbuch der Rorschach-Psychodiagnostik*. Bern: Huber.
- Borkenau, P., Friedel, A. & Wolfradt, U. (2011). Standardisierte Persönlichkeitsfragebögen. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Persönlichkeitsdiagnostik. Enzyklopädie der Psychologie B/II/4* (S. 1-72). Göttingen: Hogrefe.
- Bornstein, R. F., Rossner, S. C., Hill, E. L. & Stepanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment*, 63, 363-386.
- Bosnjak, M. & Batinic, B. (2000). Understanding the willingness to participate in online-surveys – the case of e-mail questionnaires. In B. Batinic, U.-D. Reips, M. Bosnjak & A. Werner (Eds.), *Online Social Sciences* (pp. 105-116). Zugriff unter <http://mres.gmu.edu/pmwiki/uploads/Main/oss-book.pdf>.
- Briken, P., Rettenberger, M. & Dekker, A. (2013). Was sagen „objektive“ Messverfahren über Sexualstraftäter? *Forensische Psychiatrie, Psychologie, Kriminologie*, 7, 28-33.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Cima, M., Hollnack, S., Kremer, K., Knauer, E. & Merckelbach, H. (2003a). Untersuchungen zur Simulation, Dissimulation und Supernormalität bei forensisch-psychiatrischen Patienten. In M. Osterheider (Hrsg.), *Forensik 2002. Wie sicher kann Prognose sein? 17. Eickelborner Fachtagung* (S. 112-123). Dortmund: PsychoGen.
- Cima, M., Hollnack, S., Kremer, K., Knauer, E., Schellbach-Matties, R., Klein, B. & Merckelbach, H. (2003b). Strukturierter Fragebogen Simulierter Symptome. Die deutsche Version des "Structured Inventory of Malingered Symptomatology: SIMS". *Nervenarzt*, 74, 977-986.
- Cima, M., Merckelbach, H., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R. & Muris, P. (2003c). The other side of malingering: supernormality. *The Clinical Neuropsychologist*, 17, 235-243.
- Cima, M., Pantus, M. & Dams, L. (2009). Simulation und Dissimulation in Abhängigkeit vom strafrechtlichen Kontext und der Persönlichkeit. In T. Merten & H. Dettenborn (Hrsg.), *Diagnostik der Beschwerdenvalidität* (S. 54-69). Berlin: Deutscher Psychologen Verlag.

- Cima, M., van Bergen, S. & Kremer, K. (2008). Development of the Supernormality Scale-Revised and its relationship with psychopathy. *Journal of Forensic Sciences*, 53, 975-981.
- Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Dettenborn, H. (2009). Dissimulation in der Familienrechts-Begutachtung – eine negierte Gefahr. In T. Merten & H. Dettenborn (Hrsg.), *Diagnostik der Beschwerdenvalidität* (S. 70-78). Berlin: Deutscher Psychologen Verlag.
- Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (Hrsg.). (2009). *S2 Praxisleitlinien in Psychiatrie und Psychotherapie, Band 1 Behandlungsleitlinie Persönlichkeitsstörungen*. Darmstadt: Steinkopff.
- Dilling, H., Mombour, W., Schmidt, M. H. & Schulte-Markwort, E. (2011). *Weltgesundheitsorganisation. Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F). Klinisch-diagnostische Leitlinien*. Bern: Huber.
- Dressing, H., Foerster, K., Widder, B., Schneider, F. & Falkai, P. (2011). *Stellungnahme der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) zur Anwendung von Beschwerdenvalidierungstests in der psychiatrischen Begutachtung*. Zugriff unter: http://www.dgppn.de/fileadmin/user_upload/_medien/download/pdf/stellungnahmen/2011/stn-2011-01-28-beschwerdevalidierung.pdf
- Edens, J. F. (2004). Effect of response distortion on the assessment of divergent facets of psychopathy. *Assessment*, 11, 109-112.
- Edens, J. F., Buffington, J. K. & Tomicic, T. L. (2000). An investigation of the relationship between psychopathic traits and malingering on the Psychopathic Personality Inventory. *Assessment*, 7, 281-296.
- Edens, J. F., Buffington, J. K., Tomicic, T. L. & Riley, B. D. (2001). Effects of positive impression management on the Psychopathic Personality Inventory. *Law and Human Behavior*, 25, 235-256.
- Eisenbarth, H. & Alpers, G. W. (2007). Validierung der deutschen Übersetzung des Psychopathy Personality Inventory (PPI). *Zeitschrift für Klinische Psychologie und Psychotherapie*, 36, 216-224.
- Engel, R. R. (2003). *MMPI-2. Minnesota Multiphasic Personality Inventory-2. Manual*. Bern: Huber.
- Fahrenberg, J., Hampel, R. & Selg, H. (2010). *FPI-R. Freiburger Persönlichkeitsinventar. Manual*. Göttingen: Hogrefe.

- Fiduccia, C. E. (2011). *Effects of defensiveness on the reporting of personality disorder symptoms* (Master's thesis). University of North-Texas, Denton. Zugriff unter http://digital.library.unt.edu/ark:/67531/metadc103313/m2/1/high_res_d/thesis.pdf
- Gacono, C. B., Meloy, J. R., Sheppard, K., Speth, E. & Roske, A. (1995). A clinical investigation of malingering and psychopathy in hospitalized insanity acquittees. *Journal of the American Academy of Psychiatry and the Law Online*, 23, 387-397.
- Gerger, H., Kley, H., Böhner, G. & Siebler, F. (2007). The Acceptance of Modern Myths about Sexual Aggression Scale: Development and validation in German and English. *Aggressive Behavior*, 33, 422-440.
- Goldhammer, F. & Hartig, J. (2012). Interpretation von Testresultaten und Testeichung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 173-201). Berlin: Springer.
- Graybill, D. & Heuvelman, L. R. (1993). Validity of the Children's Picture-Frustration Study: A social-cognitive perspective. *Journal of Personality Assessment*, 60, 379.
- Greene, R. L. (2007). Forensic Applications of the Minnesota Multiphasic Personality Inventory-2. In A. M. Goldstein (Ed.), *Forensic psychology: Emerging topics and expanding roles* (pp. 73-96). Hoboken, NJ: John Wiley & Sons.
- Greene, R. L. (2008). Malingering and defensiveness on the MMPI-2. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 159-181). New York: Guilford Press.
- Griffith, R. L. & Peterson, M. H. (2006). *A closer examination of applicant faking behavior*. Greenwich: Information Age Publishing.
- Grillo, J., Brown, R. S., Hilsabeck, R., Price, J. R. & Lees-Haley, P. R. (1994). Raising doubts about claims of malingering: Implications of relationships between MCMI-II and MMPI-2 performances. *Journal of Clinical Psychology*, 50, 651-655.
- Grossman, L. S. & Wasylw, O. E. (1988). A psychometric study of stereotypes: Assessment of malingering in a criminal forensic group. *Journal of Personality Assessment*, 52, 549.
- Grossman, L. S., Wasylw, O. E., Benn, A. F. & Gyoerkoe, K. L. (2002). Can sex offenders who minimize on the MMPI conceal psychopathology on the Rorschach? *Journal of Personality Assessment*, 78, 484-501.
- Hampel, R. & Selg, H. (1998). *FAF. Fragebogen zur Erfassung von Aggressivitätsfaktoren. Handanweisung*. Göttingen: Hogrefe.

- Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised (PCL-R)*. Toronto: Multi-Health Systems.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143-171). Berlin: Springer.
- Hathaway, S. R. & McKinley, J. C. (1951). *Minnesota Multiphasic Personality Inventory. Revised Edition*. New York: Psychological Corporation [dt. Adaptation Spreen, O. (Hrsg.) (1963). MMPI-Saarbrücken. Handbuch zur deutschen Ausgabe des MMPI von S. R. Hathaway und J. C. McKinley. Bern: Huber].
- Hathaway, S. R. & McKinley, J. C. (1989). *MMPI-2. Minnesota Multiphasic Personality Inventory-2*. Minneapolis: University of Minnesota [dt. Adaptation Engel, R. (Hrsg.) (2000). MMPI-2. Minnesota Multiphasic Personality Inventory-2. Bern: Huber].
- Haywood, T. W., Grossman, L. S. & Hardy, D. W. (1993). Denial and social desirability in clinical examinations of alleged sex offenders. *Journal of Nervous and Mental Disease*, 181, 183-188.
- Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior*, 16, 257-272.
- Heilbrun, K., Bennett, W. S., White, A. J. & Kelly, J. (1990). An MMPI-based empirical model of malingering and deception. *Behavioral Sciences & the Law*, 8, 45-53.
- Heilbrun, K., Marczyk, G., DeMatteo, D. & Mack-Allen, J. (2007). A principles-based approach to forensic mental health assessment: Utility and update. In A. M. Goldstein (Ed.), *Forensic psychology: Emerging topics and expanding roles*. (pp. 45-72). Hoboken, NJ: John Wiley & Sons.
- Hemphill, J. F. & Hart, S. D. (2003). Forensic and clinical issues in the assessment of psychopathy. In A. M. Goldstein & I. B. Weiner (Ed.), *Handbook of Psychology, Volume 11: Forensic Psychology* (pp. 87-107). New York: John Wiley & Sons.
- Henkel, Y. (2005). *Klinischer Validitätsvergleich zweier Selbstbeurteilungsfragebögen unter besonderer Berücksichtigung der Borderline-Persönlichkeitsstörung* (Dissertation). Universität Hamburg. Zugriff unter: <http://ediss.sub.uni-hamburg.de/volltexte/2006/3098/pdf/Dissertation.pdf>
- Herzberg, P. Y. (2011). Selbstdarstellung in Persönlichkeitsfragebögen: Das Phänomen der sozialen Erwünschtheit. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Persönlichkeitsdiagnostik. Enzyklopädie der Psychologie B/II/4* (S. 121-155). Göttingen: Hogrefe.

- Jacobson, N. S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Kaufmann, E. & Reips, U.-D. (2008). Internet-basierte Messung sozialer Erwünschtheit. Saarbrücken: VDM.
- King, J. C. (1972). *A comparison of normal and delinquent adolescents using the Rosenzweig Picture-Frustration Study* (Master's thesis). Texas Tech University, Lubbock. Zugriff unter: <http://repositories.tdl.org/ttu-ir/bitstream/handle/2346/11549/31295004441191.pdf?sequence=1>
- Kröber, H.-L. (2005). Psychologische und psychiatrische Begutachtung im Strafrecht. In H.-L. Kröber & M. Steller (Hrsg.), *Psychologische Begutachtung im Strafverfahren. Indikationen, Methoden und Qualitätsstandards* (S. 205-219). Darmstadt: Steinkopff.
- Kubinger, K. D. (2006). Ein Update der Definition von Objektiven Persönlichkeitstests: Experimentalpsychologische Verhaltensdiagnostik. In T. M. Ortner, R. T. Proyer & K. D. Kubinger (Hrsg.), *Theorie und Praxis Objektiver Persönlichkeitstests* (S. 38-52). Bern: Huber.
- Kucharski, L. T., Duncan, S., Egan, S. S. & Falkenbach, D. M. (2006). Psychopathy and malingering of psychiatric disorder in criminal defendants. *Behavioral Sciences & the Law*, 24, 633-644.
- Kuhl, J. & Kazén, M. (2009). *Persönlichkeits-Stil-und-Störungs-Inventar. PSSI. Manual*. Göttingen: Hogrefe.
- Kunzl, F. K. (2011). *Qualitätsanalyse österreichischer Gutachten zur Zurechnungsfähigkeit und Gefährlichkeitsprognose von Sexualstraftätern* (Dissertation). Universität Ulm. Zugriff unter: http://vts.uni-ulm.de/docs/2012/7828/vts_7828_11310.pdf
- Kury, H. (1987). *Die Behandlung Straffälliger. Teilband 2: Ergebnisse einer empirischen Untersuchung zum Behandlungserfolg bei jugendlichen und heranwachsenden Untersuchungshäftlingen*. Berlin: Duncker & Humblot.
- Lanyon, R. I. & Cunningham, K. S. (2005). Construct validity of the misrepresentation scales of the Psychological Screening Inventory. *Journal of Personality Assessment*, 85, 197-206.
- Leiner, D. J. (2012). *SoSci Survey* (Version 2.3.04). Zugriff unter: www.soscisurvey.de.
- Lemm, C. (2000). *Die strafrechtliche Verantwortlichkeit jugendlicher Rechtsbrecher*. Münster: Waxmann.

- Lilienfeld, S. O. & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment*, 66, 488-524.
- Lilienfeld, S. O., Wood, J. M. & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1, 27-66.
- Littmann, E. (2007). Stand und Stellenwert psychodiagnostischer Verfahren im Rahmen der Diagnostik von Persönlichkeitsstörungen in foro. In M. Lammel, W. Felber, S. Sutarski & S. Lauf (Hrsg.), *Forensische Begutachtung bei Persönlichkeitsstörungen* (S. 13-56). Berlin: MWV.
- MacNeil, B. M. & Holden, R. R. (2006). Psychopathy and the detection of faking on self-report inventories of personality. *Personality and Individual Differences*, 41, 641-651.
- Mader, E. M. (2014). *Schilderung von Beschwerden durch psychiatrische Patienten und sozialmedizinische Begutachtungsprobanden: Antwortverzerrung oder Symptombelastung?* (Dissertation). Ludwig-Maximilians-Universität München.
- Maß, R. (2001). *ESI. Eppendorfer Schizophrenie-Inventar. Manual*. Göttingen: Hogrefe.
- Maurer, M. & Jandura, O. (2009). Masse statt Klasse? Einige kritische Anmerkungen zu Repräsentativität und Validität von Online-Befragungen. In N. Jakob, H. Schoen & T. Zerback (Hrsg.), *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung* (S. 61-73). Wiesbaden: VS.
- McCrae, R. R. & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882-888.
- McEwan, T. E., Davis, M. R., MacKenzie, R. & Mullen, P. E. (2009). The effects of social desirability response bias on STAXI-2 profiles in a clinical forensic sample. *British Journal of Clinical Psychology*, 48, 431-436.
- Merten, T., Stevens, A. & Blaskewitz, N. (2009). Beschwerdevalidität und Begutachtung: eine Einführung. In T. Merten & H. Dettenborn (Hrsg.), *Diagnostik der Beschwerdevalidität* (S. 9-35). Berlin: Deutscher Psychologen Verlag.
- Millon, T. (1992). Millon Clinical Multiaxial Inventory: I & II. *Journal of Counseling & Development*, 70, 421-426.
- Merten, T. (2014). *Beschwerdenvalidierung*. Göttingen: Hogrefe.
- Mombour, W., Zaudig, M., Berger, P., Gutierrez, K., Berner, W., Berger, K., von Cranach, M., Giglhuber, O. & von Bose, M. (1996). *International Personality Disorder Examination. IPDE. ICD-10 Modul von A. W. Loranger*. Bern: Huber.

- Morrison, A. P., Wells, A. & Nothard, S. (2002). Cognitive and emotional predictors of predisposition to hallucinations in non-patients. *British Journal of Clinical Psychology*, 41, 259-270.
- Mueller-Hanson, R. A., Heggstad, E. D. & Thornton III, G. C. (2006). Individual differences in impression management: An exploration of the psychological processes underlying faking. *Psychology Science*, 48, 288-312.
- Musch, J. & Reips, U.-D. (2000). A brief history of web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. 61-88). San Diego: Academic Press.
- Myerholtz, L. E. & Rosenberg, H. (1997). Screening DUI offenders for alcohol problems: Psychometric assessment of the Substance Abuse Subtle Screening Inventory. *Psychology of Addictive Behaviors*, 11, 155-165.
- Murray, H. A. (1991). *Thematic Apperception Test*. Cambridge: Harvard University Press.
- Nedopil, N., Tippelt, S. & Stübner, S. (2012). *Zwischenbericht zu Projektteil A der Begleituntersuchung der psychotherapeutischen Fachambulanz für Sexualstraftäter in Würzburg* (Unveröffentlichter Bericht). Ludwig-Maximilians-Universität München.
- Nedopil, N., Tippelt, S. & Stübner, S. (2013). *Endbericht zu Projektteil A der Begleituntersuchung der psychotherapeutischen Fachambulanz für Sexualstraftäter in Würzburg* (Unveröffentlichter Bericht). Ludwig-Maximilians-Universität München.
- Nicholson, R. A. & Hogan, R. (1990). The construct validity of social desirability. *American Psychologist*, 45, 290-292.
- Ones, D. S., Viswesvaran, C. & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660-679.
- Ostendorf, F. & Angleitner, A. (2004). *NEO-PI-R. NEO Persönlichkeitsinventar nach Costa und McCrae. Revidierte Fassung*. Göttingen: Hogrefe.
- Otto, R. K. (2008). Challenges and advances in assessment of response style in forensic examination contexts. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 365-375). New York: Guilford Press.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.

- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and psychological attitudes* (pp. 17-59). San Diego: Academic Press.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah: Erlbaum.
- Pauls, C. A. & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences*, 35, 263-275.
- Poythress, N. G., Edens, J. F. & Lilienfeld, S. O. (1998). Criterion-related validity of the Psychopathic Personality Inventory in a prison sample. *Psychological Assessment*, 10, 426-430.
- Poythress, N. G., Edens, J. F. & Watkins, M. M. (2001). The relationship between psychopathic personality features and malingering symptoms of major mental illness. *Law and Human Behavior*, 25, 567-582.
- Rambow, J., Elsner, K., Feelgood, S. & Hoyer, J. (2008). Einstellungen zum Kindesmissbrauch. *Zeitschrift für Sexualforschung*, 21, 341,355.
- Rauchfleisch, U. (1979a). *Handbuch zum Rosenzweig Picture-Frustration Test (PFT). Band 1: Grundlagen, bisherige Resultate und Anwendungsmöglichkeiten des PFT*. Bern: Huber.
- Rauchfleisch, U. (1979b). *Handbuch zum Rosenzweig Picture-Frustration Test (PFT). Band 2: Manual zur Durchführung des PFT und Neueichung der Testformen für Kinder und Erwachsene*. Bern: Huber.
- Ray, J. V., Hall, J., Rivera-Hudson, N., Poythress, N. G., Lilienfeld, S. O. & Morano, M. (2013). The relation between self-reported psychopathic traits and distorted response styles: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment*, 4, 1-14.
- Revers, W. J. (1993). *Der thematische Apperzeptionstest (TAT). Handbuch zur Verwendung des TAT in der psychologischen Persönlichkeitsdiagnostik*. Bern: Huber.
- Rogers, R. (2008a). An introduction to response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 3-13). New York: Guilford Press.
- Rogers, R. (2008b). Detection strategies for malingering and deception. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 14-38). New York: Guilford Press.

- Rogers, R. (2008c). Current status of clinical methods. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 391-410). New York: Guilford Press.
- Rogers, R. (2008d). Researching response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 411-434). New York: Guilford Press.
- Rorschach, H. (1994). *Rorschach-Test. Psychodiagnostik Tafeln*. Bern: Huber.
- Rosenzweig, S. (1976). Aggressive Behavior and the Rosenzweig Picture-Frustration (P-F) Study. *Journal of Clinical Psychology*, 32, 885-891.
- Saß, H., Wittchen, H.-U., Zaudig, M. & Houben, I. (2003). *Diagnostisches und Statistisches Manual Psychischer Störungen – Textrevision – DSM-IV-TR*. Göttingen: Hogrefe.
- Scheurer, H. & Richter, P. (2005). Psychologische Persönlichkeitsdiagnostik: Zur Bedeutung von Persönlichkeitsfragebogen bei der Begutachtung der Schuldfähigkeit. In H.-L. Kröber & M. Steller (Hrsg.), *Psychologische Begutachtung im Strafverfahren* (S. 39-60). Darmstadt: Steinkopff.
- Schmand, B. & Merten, T. (2009). Grenzen und Probleme der Beschwerdenvalidierung. In T. Merten & H. Dettenborn (Hrsg.), *Diagnostik der Beschwerdenvalidität* (S. 230-245). Berlin: Deutscher Psychologen Verlag.
- Schmidt, K.-H. & Metzler, P. (1992). *WST. Wortschatztest*. Weinheim: Beltz Test.
- Schneewind, K. A. & Graf, J. (1998). *Der 16-Persönlichkeits-Faktoren-Test. Revidierte Fassung. 16 PF-R – deutsche Ausgabe des 16 PF Fifth Edition – Testmanual*. Bern: Huber.
- Schröder, K. (2004). *Diagnostische Methoden im Vergleich bei angstnahen Persönlichkeitsstörungen* (Dissertation). Universität Hamburg. Zugriff unter: http://ediss.sub.uni-hamburg.de/volltexte/2005/2529/pdf/Katrin_Schroeder.pdf
- Sellbom, M. & Bagby, R. M. (2008). Response style on multiscale inventories. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 182-206). New York: Guilford Press.
- Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T. & Duncan, S. (2010). Utility of the MMPI-2-RF (Restructured Form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment*, 22, 22-31.
- Sewell, K. W. (2008). Dissimulation on projective measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 207-217). New York: Guilford Press.

- Smith, G. P. & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law Online*, 25, 183-189.
- Smith, P. & Waterman, M. (2004). Role of experience in processing bias for aggressive words in forensic and non-forensic populations. *Aggressive Behavior*, 30, 105-122.
- Steffan, J. S., Kroner, D. G. & Morgan, R. D. (2007). Effect of symptom information and intelligence in dissimulation: An examination of faking response styles by inmates on the Basic Personality Inventory. *Assessment*, 14, 22-34.
- Steinmayr, R. & Amelang, M. (2011). Zur Diagnostik konnotativ negativ besetzter Konstrukte. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Persönlichkeitsdiagnostik. Enzyklopädie der Psychologie B/II/4* (S. 639-688). Göttingen: Hogrefe.
- Steller, M. (1994). Diagnostischer Prozess. In R. D. Stieglitz & U. Baumann (Hrsg.), *Psychodiagnostik psychischer Störungen* (S. 37-46). Stuttgart: Enke.
- Taddicken, M. (2009). Die Bedeutung von Methodeneffekten der Online-Befragung: Zusammenhänge zwischen computer-vermittelter Kommunikation und erreichbarer Datengüte. In N. Jakob, H. Schoen & T. Zerback (Hrsg.), *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung* (S. 91-107). Wiesbaden: VS.
- Tan, L. & Grace, R. C. (2008). Social desirability and sexual offenders: A review. *Sexual Abuse: Journal of Research and Treatment*, 20, 61-87.
- Tippelt, S., Yundina, E., Nedopil, N. & Veress, K. (2010, Oktober). *Erfassung von Antwortverzerrungen im forensischen Kontext – Supernormalität bei Schuldfähigkeits- und Prognosebegutachtungen*. Vortrag auf der 25. Münchner Herbsttagung der Arbeitsgemeinschaft für Methodik und Dokumentation in der Forensischen Psychiatrie (AGFP), München.
- Vane, J. R. (1954). Implications of the performance of delinquent girls on the Rosenzweig Picture-Frustration Study. *Journal of Consulting Psychology*, 18, 414-414.
- Veress, K. (2010). *Faking good und Faking bad: Über Antwortverhalten im gutachterlichen Kontext. Eine explorative Studie zu Antwortstilen* (Unveröffentlichte Diplomarbeit). Ludwig-Maximilians-Universität München.
- Veress, K., Merten, T. & Yundina, E. (2010). *Supernormalität-Skala. Autorisierte deutsche Version der Supernormality Scale – Revised von M. Cima*. Unveröffentlichtes Manuskript, Ludwig-Maximilians-Universität München.

- Viswesvaran, C. & Ones, D. S. (1999). Meta-Analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197-210.
- Walters, G. D. (1988). Assessing dissimulation and denial on the MMPI in a sample of maximum security, male inmates. *Journal of Personality Assessment*, 52, 465-474.
- Weiner, I. B. (2007). Rorschach assessment in forensic cases. In A. M. Goldstein (Ed.), *Forensic psychology. Emerging topics and expanding roles* (pp. 127–153). Hoboken, NJ: John Wiley & Sons.
- Weisband, S. & Kiesler, S. (1996). Self disclosure on computer forms: meta-analysis and implications. Zugriff unter: http://www.sigchi.org/chi96/proceedings/papers/Weisband/sw_txt.htm
- Werner, R. (2005). *Implizite Aggressionsneigung: Messung und Verhaltensvorhersage* (Unveröffentlichte Dissertation). Universität Leipzig.
- Wetzler, S. & Marlowe, D. (1990). “Faking bad” on the MMPI, MMPI-2, and Millon-II. *Psychological Reports*, 67, 1117-1118.
- Whyte, S., Fox, S. & Coxell, A. (2006). Reporting of personality disorder symptoms in a forensic inpatient sample: Effects of mode of assessment and response style. *Journal of Forensic Psychiatry & Psychology*, 17, 431-441.
- Widder, B. (2007). Beurteilung der Beschwerdenuvalidität. In B. Widder & P. W. Gaidzik (Hrsg.), *Begutachtung in der Neurologie* (S. 64-92). Stuttgart: Thieme.
- Wilson, R. J. (1999). Emotional Congruence in Sexual Offenders Against Children. *Sexual Abuse: A Journal of Research and Treatment*, 11, 33-47.
- Wise, E. A. (2002). Relationships of personality disorders with MMPI-2 malingering, defensiveness, and inconsistent response scales among forensic examinees. *Psychological Reports*, 90, 760-766.
- Wittkowski, J. (2011). Projektive Verfahren. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Persönlichkeitsdiagnostik. Enzyklopädie der Psychologie B/II/4* (S. 299-410). Göttingen: Hogrefe.
- Yundina, E. & Nedopil, N. (2010). Indirekte Diagnostik pädosexueller Neigungen. In J. L. Müller (Hrsg.), *Neurobiologie forensisch relevanter Störungen* (S. 453-460). Stuttgart: Kohlhammer.
- Ziegler, M., Schmidt-Atzert, L., Bühner, M. & Krumm, S. (2007). Fakability of different measurement methods for achievement motivation: Questionnaire, semi-projective, and objective. *Psychology Science*, 49, 291-307.

Zimmermann, L. C. (2012). *Response styles in forensic assessment: The relationship of supernormality and malingering with legal context, personality disorder traits and cognitive distortions* (Unpublished master's thesis). Maastricht University.

8 Anhang

Anhang A: Erhebungsinstrumente

A1: Onlinefragebogen Begrüßungstext

Die Abteilung für **Forensische Psychiatrie** der Ludwig-Maximilians-Universität München führt verschiedene Forschungsprojekte zum Antwortverhalten von Straftätern durch. Die aktuelle Untersuchung beschäftigt sich mit der **Selbstdarstellung und dem Antwortverhalten von (psychisch kranken) Straftätern bei Begutachtungssituationen zur Legalprognose**. Das Wissen über solche Prozesse ist deshalb von besonderer Bedeutung, weil sie die Offenheit und damit die Grundlage einer adäquaten Prognosestellung bestimmen. Damit können die Ergebnisse dieser Studie bedeutsam für die Gestaltung des weiteren Haftverlaufs bzw. der weiteren Behandlung psychisch kranker Straftäter sein. Um deren Antworttendenzen erfassen zu können, benötigen wir eine Vergleichsstichprobe aus der Allgemeinbevölkerung. Wir ersuchen Sie deshalb, den Fragebogen als Teilnehmer der Vergleichsstichprobe zu beantworten.

Sie, als Teilnehmer dieser Untersuchung werden gebeten, **Fragen aus thematisch unterschiedlichen Bereichen** zu beantworten, z.B. über allgemeine und persönliche Einstellungen, Verhaltensweisen und Emotionen.

Für die Beantwortung aller Fragen werden Sie insgesamt ca. 10 bis maximal 20 Minuten benötigen.

Die Daten werden ausschliesslich für wissenschaftliche Zwecke verwendet, die Erhebung ist anonym. Es werden weder Name noch IP-Adresse gespeichert. Damit ist auch ausgeschlossen, dass Sie in etwaigen Veröffentlichungen namentlich genannt werden. Sobald der Forschungszweck es zulässt, werden Ihre Daten gelöscht. Eine Zuordnung der erhobenen Daten zu Ihrer Person wird zu keinem Zeitpunkt möglich sein.

Bitte beachten Sie beim Ausfüllen des Fragebogens auch Folgendes:

Versuchen Sie den Fragebogen **möglichst ohne Unterbrechungen** zu bearbeiten.

Bitte lassen Sie beim Ausfüllen keine Frage aus, da sonst möglicherweise der ganze Fragebogen unbrauchbar wird. Sollten Sie sich bei einer Antwort einmal nicht sicher sein, so kreuzen Sie bitte **die am ehesten zutreffende Antwort** an.

Vielen Dank für Ihre Teilnahme!

Bei Rückfragen können Sie uns jederzeit gerne kontaktieren
(e-Mail: susanne.tippelt@med.uni-muenchen.de)

Weiter



Abteilung für Forensische Psychiatrie
Ludwig-Maximilians-Universität-München
www.forensik-muenchen.de

0% ausgefüllt

Anhang B: Formeln für die Berechnung der diagnostischen Validität

		SN-S		Σ
		unauffällig	auffällig	
Untersuchungs- gruppe	KG	a	b	a + b
	Fake-KG	c	d	c + d
	Σ	a + c	b + d	N

Sensitivität: $\frac{d}{c + d} \times 100$

Spezifität: $\frac{a}{a + b} \times 100$

positiver Vorhersagewert: $\frac{d}{b + d} \times 100$

negativer Vorhersagewert: $\frac{a}{a + c} \times 100$

Gesamttrefferquote: $\frac{a + d}{N} \times 100$

Anhang C: Tabellen und Abbildungen

Tabelle 25

Mittwerte (*M*), Standardabweichungen (*SD*) und korrigierte Trennschärfen (*r*) für den Gesamtwert der 34 SN-S-Items in der Gesamtstichprobe und den Substichproben (Fortsetzung S. 136-137)

Item	GS		SF		RP		ZB		GN ^a	KG		Fake-KG	
	<i>N</i> = 281		<i>n</i> = 33		<i>n</i> = 42		<i>n</i> = 20		<i>n</i> = 8	<i>n</i> = 114		<i>n</i> = 65	
	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>
1	1.7 (.81)	.40	1.6 (.97)	.81	2.1 (.42)	.53	2.0 (1.1)	.31	1.1 (.39)	1.9 (.81)	.25	1.6 (.70)	.16
2	2.6 (.90)	.54	2.6 (.86)	.59	3.1 (.75)	.63	3.1 (.91)	.65	1.9 (.90)	2.9 (.75)	.30	2.1 (.80)	.41
4	2.9 (.91)	.23	3.0 (.95)	.38	1.8 (.86)	.19	3.0 (.76)	.05	2.9 (.90)	3.0 (.90)	.28	2.7 (1.0)	-.03
5	2.0 (.90)	.65	2.3 (1.1)	.63	2.0 (.83)	.60	2.9 (.91)	.67	1.6 (.54)	2.2 (.80)	.53	1.5 (.59)	.39
7	1.9 (.87)	.58	2.3 (.98)	.48	1.3 (.65)	.72	2.5 (1.1)	.71	1.3 (.49)	2.0 (.81)	.37	1.4 (.55)	.39
8	1.6 (.82)	.48	1.7 (.98)	.59	3.8 (.48)	.33	1.9 (.88)	.55	1.3 (.76)	1.7 (.86)	.41	1.3 (.66)	.28
9	3.0 (1.0)	.21	3.6 (.87)	.12	2.7 (1.1)	.17	2.7 (1.1)	.43	2.9 (.69)	3.0 (.98)	.23	2.6 (1.0)	.19
11	2.5 (.99)	.39	2.7 (1.2)	.42	1.6 (1.1)	.38	2.8 (.89)	.63	2.4 (.54)	2.5 (.96)	.31	2.2 (.88)	.33
12	1.8 (.92)	.65	1.9 (.99)	.74	1.6 (.83)	.72	2.5 (1.1)	.77	1.4 (.79)	2.0 (.90)	.55	1.4 (.70)	.20
13	2.1 (1.0)	.48	2.3 (1.2)	.71	1.8 (.82)	.45	2.9 (1.1)	.60	2.0 (.82)	2.3 (.94)	.29	1.9 (.89)	.20
15	2.0 (.90)	.66	2.2 (1.1)	.73	1.4 (.82)	.60	3.0 (1.0)	.71	1.7 (.95)	2.0 (.82)	.53	1.5 (.64)	.51
16	1.4 (.65)	.34	1.5 (.87)	.23	1.5 (.62)	.49	1.4 (.50)	.30	1.0 (0.0)	1.5 (.69)	.35	1.2 (.51)	.20
18	1.6 (.82)	.48	1.9 (1.1)	.59	1.3 (.77)	.37	1.9 (1.0)	.50	1.3 (4.9)	1.6 (.78)	.37	1.3 (.63)	.29

Fortsetzung Tabelle 25

	GS		SF		RP		ZB		GN ^a	KG		Fake-KG	
	<i>N</i> = 281		<i>n</i> = 33		<i>n</i> = 42		<i>n</i> = 20		<i>n</i> = 8	<i>n</i> = 114		<i>n</i> = 65	
Item	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>
19	1.5 (.76)	.58	1.8 (1.1)	.89	1.7 (.53)	.21	1.9 (1.1)	.67	1.0 (0.0)	1.6 (.75)	.33	1.2 (.40)	.39
21	1.8 (.98)	.52	2.0 (1.2)	.73	1.4 (.92)	.31	3.0 (1.1)	.78	1.0 (0.0)	1.7 (.88)	.33	1.4 (.77)	.10
22	1.4 (.74)	.62	1.7 (.92)	.67	2.1 (.70)	.64	1.9 (1.0)	.76	1.4 (.79)	1.5 (.76)	.48	1.0 (.17)	.17
23	2.2 (.88)	.63	2.6 (1.1)	.66	1.4 (.78)	.62	3.4 (.82)	.80	2.0 (1.0)	2.2 (.77)	.44	1.8 (.59)	.36
25	1.6 (.77)	.48	1.9 (1.0)	.65	1.4 (.63)	.45	1.7 (.80)	.62	1.6 (.79)	1.6 (.79)	.41	1.4 (.57)	.23
26	1.6 (.76)	.48	1.9 (1.0)	.34	1.3 (.70)	.43	2.0 (.89)	.48	1.1 (.34)	1.7 (.74)	.39	1.2 (.46)	.20
28	1.3 (.65)	.46	1.9 (1.1)	.39	1.1 (.60)	.43	1.7 (.92)	.42	1.0 (0.0)	1.3 (.48)	.32	1.1 (.24)	.28
29	1.1 (.41)	.42	1.4 (.78)	.60	1.6 (.26)	.20	1.2 (.41)	.42	1.0 (0.0)	1.1 (.39)	.28	1.0 (.12)	.26
31	1.7 (.84)	.45	1.9 (1.0)	.33	1.9 (.86)	.41	2.2 (1.1)	.44	1.9 (.90)	1.8 (.79)	.43	1.3 (.50)	-.01
32	1.8 (.93)	.44	2.0 (1.2)	.55	1.8 (1.0)	.36	2.2 (1.0)	.60	1.9 (1.2)	1.8 (.90)	.29	1.4 (.61)	.32
33	2.0 (.97)	.63	2.3 (1.1)	.69	1.3 (.95)	.69	2.9 (.99)	.50	1.1 (.38)	2.1 (.97)	.53	1.6 (.65)	.36
35	1.4 (.71)	.42	1.8 (1.0)	.66	2.1 (.56)	.35	1.7 (.93)	.45	1.0 (0.0)	1.3 (.65)	.17	1.2 (.52)	.08
36	2.1 (1.1)	.62	2.6 (1.3)	.76	1.3 (.94)	.63	3.0 (1.1)	.77	2.4 (1.1)	2.1 (1.0)	.49	1.5 (.71)	.11
38	1.3 (.59)	.44	1.6 (1.1)	.67	2.0 (.54)	.40	1.4 (.60)	.33	1.1 (.38)	1.3 (.48)	.23	1.1 (.27)	.07
39	2.2 (.98)	.66	2.5 (1.3)	.68	1.1 (.94)	.65	3.4 (.88)	.75	1.4 (.54)	2.5 (.92)	.52	1.7 (.66)	.34
40	1.5 (.82)	.49	1.7 (1.0)	.81	2.1 (.42)	.41	1.9 (1.1)	.48	1.1 (.39)	1.7 (.89)	.33	1.3 (.57)	.28

Fortsetzung Tabelle 25

	GS		SF		RP		ZB		GN ^a		KG		Fake-KG	
	<i>N</i> = 281		<i>n</i> = 33		<i>n</i> = 42		<i>n</i> = 20		<i>n</i> = 8		<i>n</i> = 114		<i>n</i> = 65	
Item	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	
42	2.2 (.98)	.38	2.2 (1.2)	.31	2.2 (.87)	.48	2.4 (1.1)	.56	1.9 (.90)	2.5 (1.0)	.24	1.9 (.78)	.35	
44	2.5 (.90)	.44	2.9 (.93)	.55	1.8 (.88)	.77	2.6 (1.0)	.44	1.9 (.90)	2.6 (.86)	.25	2.3 (.79)	.32	
46	2.0 (.97)	.44	1.9 (.93)	.52	1.4 (.81)	.23	2.6 (1.1)	.74	1.9 (1.2)	2.2 (1.0)	.25	1.6 (.71)	.35	
48	1.51 (.78)	.61	2.0 (.95)	.50	1.8 (.58)	.48	1.9 (1.1)	.78	1.4 (.79)	1.6 (.78)	.57	1.1 (.42)	.09	
49	1.91 (.90)	.52	2.0 (1.1)	.53	2.1 (.77)	.67	2.4 (1.0)	.55	1.6 (.98)	2.2 (.87)	.34	1.4 (.55)	.16	

Anmerkung: ^aIn dieser Stichprobe konnten wegen fehlender Varianz bei sechs Items keine Trennschärfen für den Gesamtwert berechnet werden.

Tabelle 26

Mittwerte (M), Standardabweichungen (SD) und korrigierte Trennschärfen (r) für den Gesamtwert der 23 Items der PPI-Skala Unaufrichtige Beantwortung in der Gesamtstichprobe und den Substichproben (Fortsetzung S. 139)

Item	GS		SF		RP ^a	KG		Fake-KG ^b
	N = 221		n = 19		n = 22	n = 117		n = 63
	M (SD)	r	M (SD)	r	M (SD)	M (SD)	r	M (SD)
1	1.0 (.21)	.08	1.2 (.38)	.27	1.0 (0.0)	1.0 (.24)	.09	1.0 (0.0)
2	1.9 (.87)	.21	2.0 (.85)	.09	2.5 (1.2)	1.8 (.79)	.20	1.8 (.81)
3	1.3 (.72)	-.01	1.5 (.77)	.01	1.2 (.66)	1.2 (.59)	-.12	1.4 (.91)
4	2.6 (.97)	.24	2.7 (1.0)	.31	2.4 (1.3)	2.4 (.87)	.18	2.9 (.93)
5	1.9 (.83)	.41	1.9 (.81)	.00	2.3 (1.1)	1.7 (.75)	.39	2.2 (.79)
6	1.1 (.34)	.20	1.3 (.75)	.12	1.1 (.29)	1.1 (.24)	.25	1.1 (.30)
7	2.5 (.84)	.28	2.6 (.96)	-.25	3.1 (.97)	2.3 (.97)	.14	2.7 (.80)
8	2.2 (.98)	.44	2.6 (.90)	.41	2.5 (1.1)	1.9 (.91)	.34	2.5 (.95)
9	1.6 (.80)	.17	1.9 (.81)	.26	1.6 (.51)	1.5 (.84)	.17	1.5 (.80)
10	2.5 (.84)	-.36	2.6 (1.1)	-.13	2.4 (.85)	2.6 (.82)	-.43	2.2 (.75)
11	2.0 (.91)	.44	2.1 (.88)	-.26	3.1 (.87)	1.7 (.80)	.33	2.1 (.81)
12	1.4 (.71)	.36	1.7 (1.0)	.28	1.6 (.96)	1.3 (.58)	.31	1.5 (.69)
13	1.2 (.50)	.20	1.7 (1.2)	.25	1.2 (.43)	1.1 (.38)	.29	1.0 (.25)
14	2.6 (.86)	.34	2.5 (.84)	.28	3.1 (.61)	2.4 (.86)	.35	2.8 (.84)
15	1.1 (.44)	.11	1.5 (.96)	-.23	1.2 (.40)	1.1 (.40)	.23	1.0 (0.0)

Fortsetzung Tabelle 26

	GS		SF		RP ^a	KG		Fake-KG ^b
	<i>N</i> = 221		<i>n</i> = 19		<i>n</i> = 22	<i>n</i> = 117		<i>n</i> = 63
Item	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>r</i>	<i>M</i> (<i>SD</i>)
16	1.9 (.82)	.44	2.2 (.83)	-.09	2.6 (.73)	1.6 (.70)	.42	2.1 (.78)
17	1.3 (.56)	.21	1.3 (.58)	.14	1.5 (.86)	1.2 (.49)	.22	1.3 (.54)
18	1.9 (.83)	.45	2.3 (.99)	.32	2.4 (.91)	1.7 (.74)	.28	2.1 (.77)
19	1.1 (.41)	.28	1.3 (.48)	.36	1.1 (.35)	1.1 (.35)	.29	1.2 (.49)
20	1.9 (.89)	-.03	2.0 (.78)	-.44	2.1 (.95)	1.8 (.93)	-.03	1.9 (.82)
21	1.6 (.88)	.36	2.1 (1.1)	.17	2.3 (1.2)	1.4 (.66)	.22	1.6 (.92)
22	2.2 (.94)	.35	2.1 (.97)	-.11	2.3 (1.0)	2.0 (.89)	.36	2.5 (.91)
23	1.2 (.54)	.20	1.6 (1.1)	.37	1.2 (.40)	1.1 (.46)	.22	1.1 (.41)

Anmerkung: ^aIn dieser Stichprobe konnten wegen fehlender Varianz bei einem Item keine Trennschärfen für den Gesamtwert berechnet werden; ^bIn dieser Stichprobe konnten wegen fehlender Varianz bei zwei Items keine Trennschärfen für den Gesamtwert berechnet werden.

Tabelle 27

Mittwerte (M), Standardabweichungen (SD) und korrigierte Trennschärfen (r) für den Gesamtwert der 11 Items der PPI-Skala Soziale Erwünschtheit in der Gesamtstichprobe und den Substichproben

Item	GS		SF		RP		KG		Fake-KG	
	N = 229		n = 19		n = 22		n = 122		n = 66	
	M (SD)	r	M (SD)	r	M (SD)	r	M (SD)	r	M (SD)	r
1	1.9 (.86)	.23	2.0 (.85)	-.04	2.5 (1.2)	.15	1.8 (.78)	.27	1.8 (.82)	.15
2	2.6 (.97)	.36	2.7 (1.0)	.45	2.4 (1.3)	.42	2.5 (.87)	.34	2.9 (.95)	.34
3	1.9 (.83)	.52	1.9 (.81)	.45	2.3 (1.1)	.20	1.7 (.75)	.47	2.2 (.80)	.59
4	2.5 (.96)	.37	2.6 (.96)	.37	3.1 (.97)	.39	2.3 (.99)	.25	2.7 (.79)	.30
5	2.2 (.97)	.49	2.6 (.90)	.59	2.5 (1.1)	.37	1.9 (.91)	.42	2.5 (.93)	.41
6	2.0 (.92)	.46	2.1 (.88)	.15	3.1 (.87)	.53	1.8 (.84)	.33	2.2 (.83)	.41
7	1.4 (.71)	.38	1.7 (1.0)	.34	1.6 (.96)	.39	1.3 (.58)	.30	1.5 (.69)	.43
8	2.6 (.86)	.41	2.5 (.84)	.38	3.1 (.61)	.45	2.4 (.87)	.38	2.8 (.82)	.20
9	1.9 (.81)	.55	2.2 (.83)	.52	2.6 (.73)	.34	1.5 (.69)	.46	2.1 (.78)	.43
10	1.9 (.83)	.46	2.3 (.99)	.37	2.4 (.91)	.36	1.7 (.73)	.30	2.1 (.76)	.51
11	2.2 (.93)	.44	2.1 (.97)	.22	2.3 (1.0)	.17	2.0 (.88)	.39	2.5 (.90)	.62

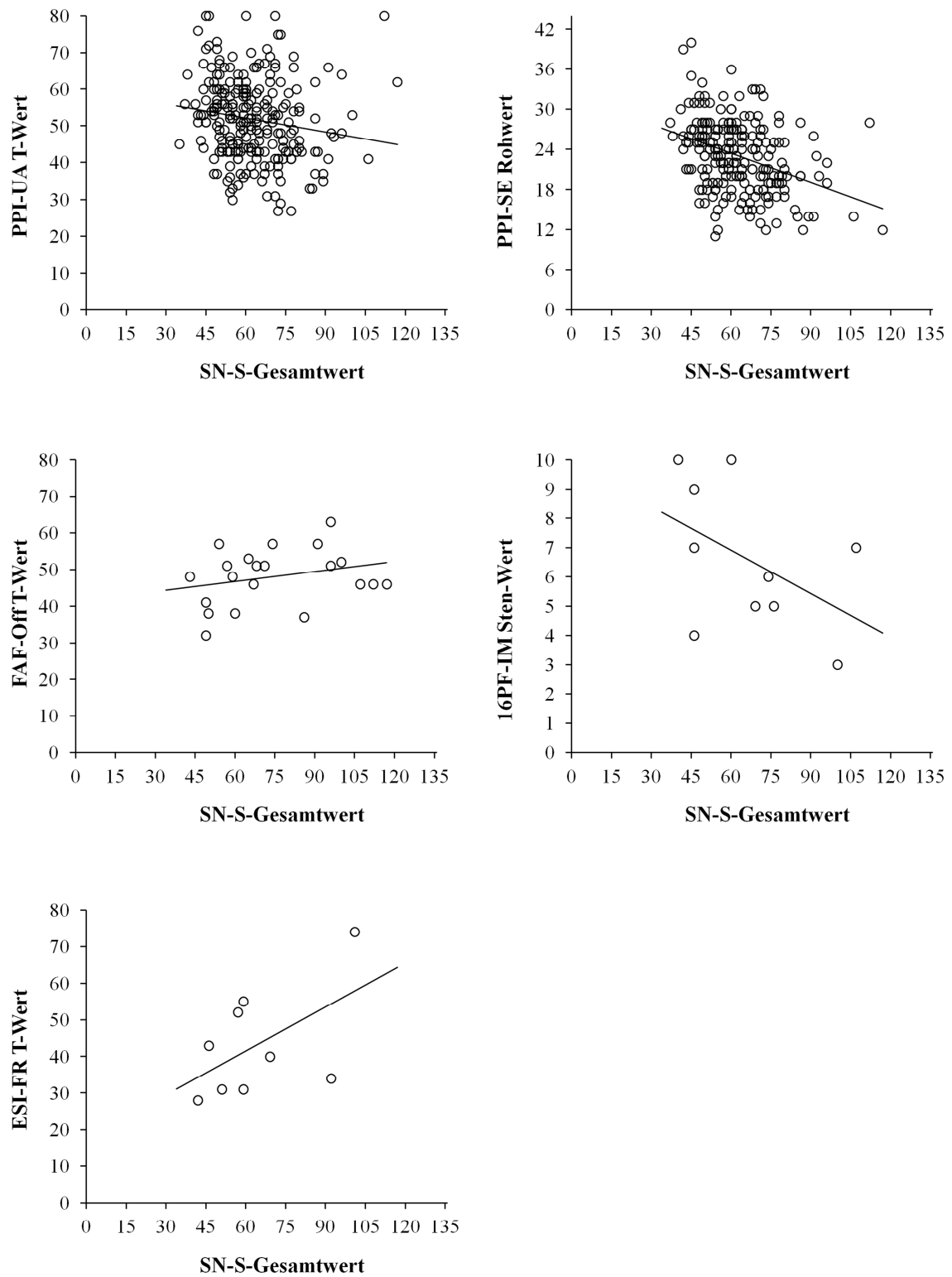


Abbildung 2. Zusammenhang zwischen dem SN-S-Gesamtwert und den Werten auf den Kontrollskalen des PPI-R, des FAF, des 16 PF-R und des ESI

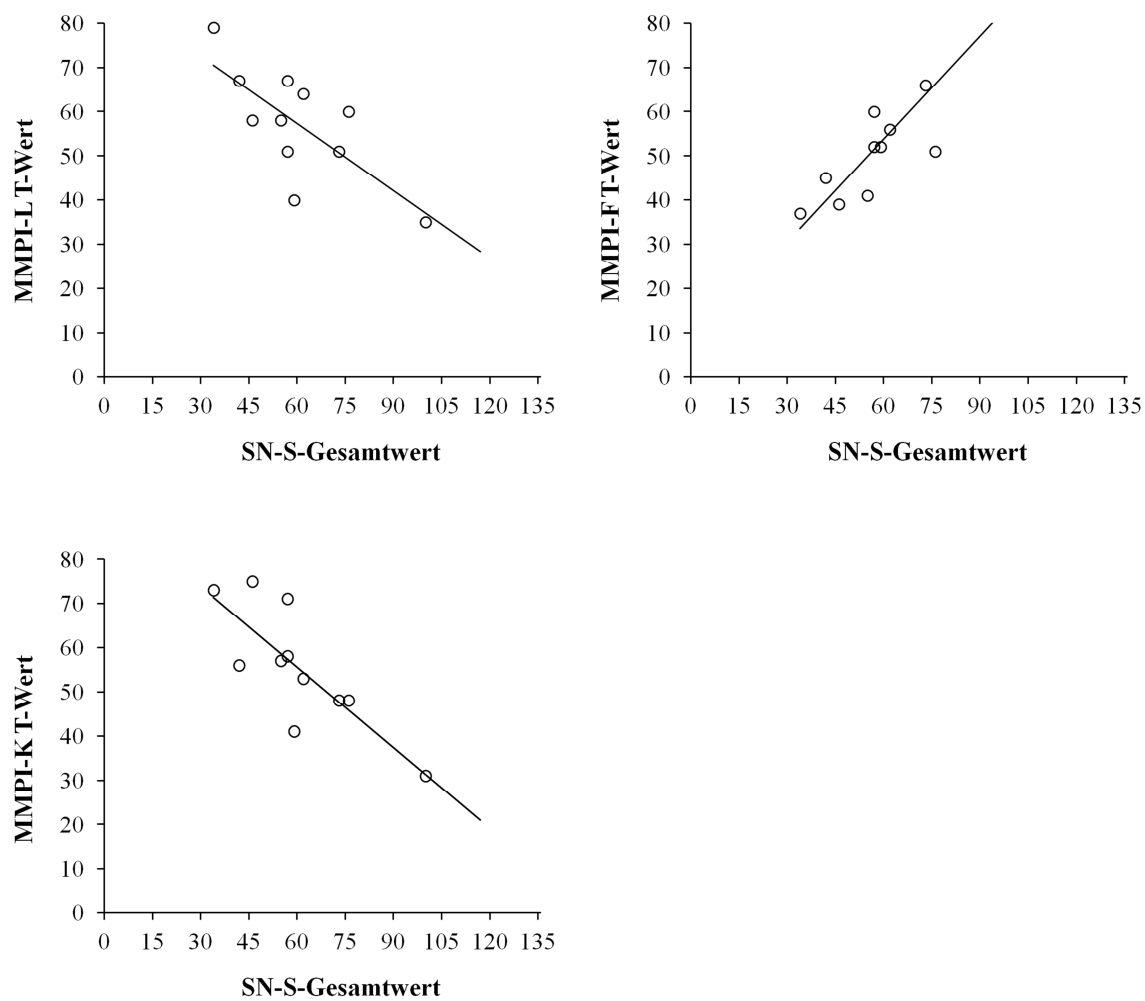


Abbildung 3. Zusammenhang zwischen dem SN-S-Gesamtwert und den Werten auf den Kontrollskalen des MMPI-2

Tabelle 28

Übereinstimmung der Klassifikation des Antwortverhaltens durch SN-S-Subskalen und die Kontrollskalen des PPI-R (Fortsetzung S. 144)

SN-S-Skala	Antwortverhalten	PPI-UA		Σ
		unauffällig	auffällig	
Zwang ^a	unauffällig	135	26	161
	supernormal	56	21	77
	Σ	191	47	238
Minimierung von Aggression ^b	unauffällig	96	95	191
	supernormal	22	25	47
	Σ	118	120	238
		PPI-SE		Σ
		unauffällig	auffällig	
Soziale Erwünschtheit ^c	unauffällig	65	31	96
	supernormal	51	81	132
	Σ	116	112	228
Affektive Symptome ^d	unauffällig	59	33	92
	supernormal	57	79	136
	Σ	116	112	228
Zwang ^e	unauffällig	86	66	152
	supernormal	30	46	76
	Σ	116	112	228
Psychotische Symptome ^f	unauffällig	45	26	71
	supernormal	71	86	157
	Σ	116	112	228

Fortsetzung Tabelle 28		PPI-SE		
SN-S-Skala	Antwortverhalten	unauffällig	auffällig	Σ
Aggression ^g	unauffällig	95	59	154
	supernormal	21	53	74
	Σ	116	112	228
Minimierung depressiver Beschwerden ^h	unauffällig	73	45	118
	supernormal	38	65	103
	Σ	111	110	221
Minimierung von Aggression ⁱ	unauffällig	65	27	92
	supernormal	50	84	134
	Σ	115	111	226
Minimierung von ungewöhnlichem Erleben ^j	unauffällig	68	49	117
	supernormal	44	63	107
	Σ	112	112	224

Anmerkung: Dargestellt sind lediglich die Skalen, bei denen sich signifikante Übereinstimmungen zeigten. ^a $\chi^2(1, N = 238) = 4.07, p < .05$; ^b $\chi^2(1, N = 236) = 6.18, p < .05$; ^c $\chi^2(1, N = 228) = 18.80, p < .001$; ^d $\chi^2(1, N = 228) = 10.84, p < .01$; ^e $\chi^2(1, N = 228) = 5.93, p < .05$; ^f $\chi^2(1, N = 228) = 6.45, p < .05$; ^g $\chi^2(1, N = 228) = 22.19, p < .001$; ^h $\chi^2(1, N = 221) = 13.72, p < .001$; ⁱ $\chi^2(1, N = 226) = 24.26, p < .001$; ^j $\chi^2(1, N = 224) = 6.46, p < .05$

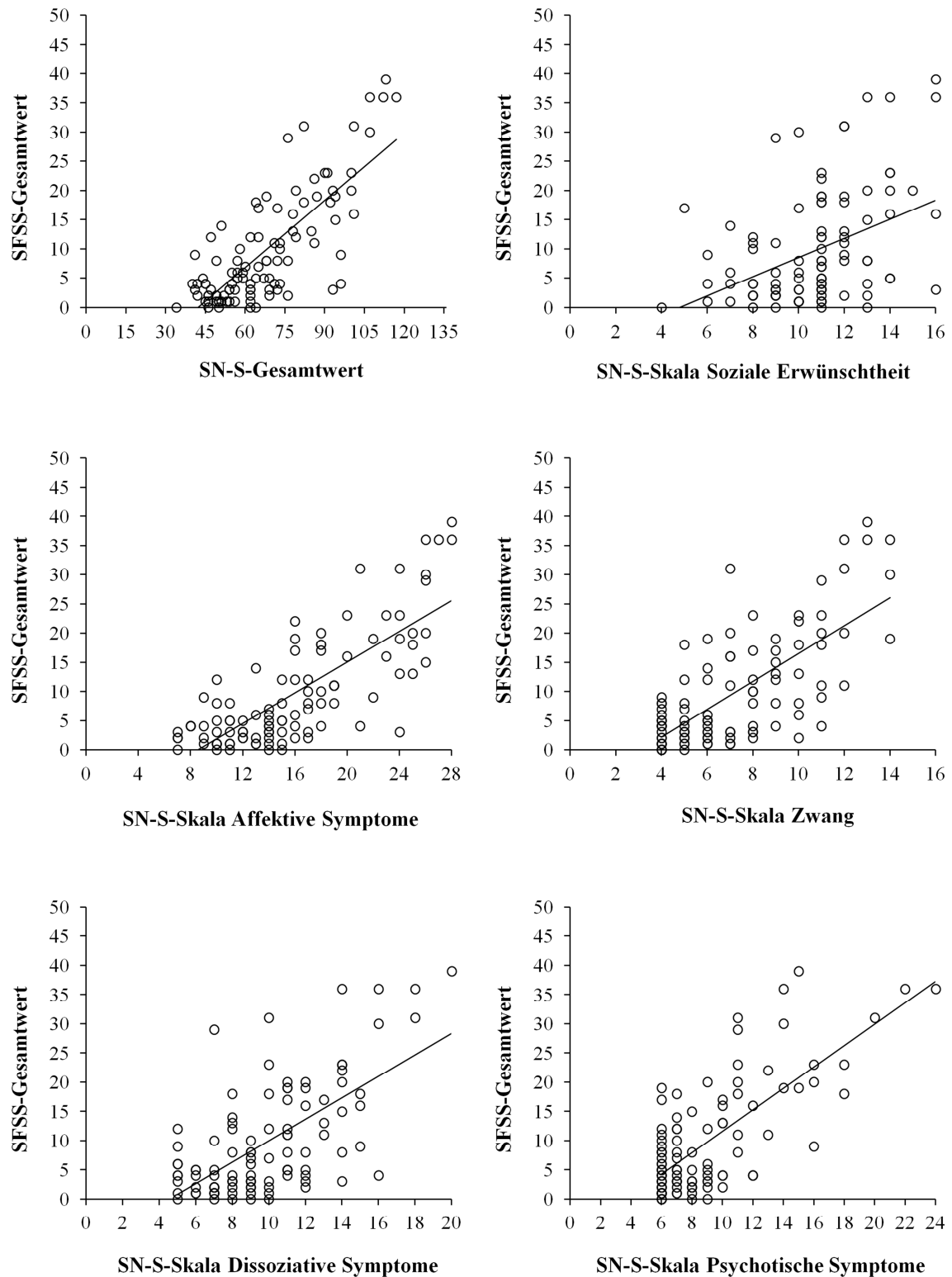


Abbildung 4. Zusammenhang zwischen dem SFSS-Gesamtwert und dem Gesamtwert der SN-S sowie den Subskalen Affektive Symptome, Zwang, Dissoziative Symptome und Psychotische Symptome

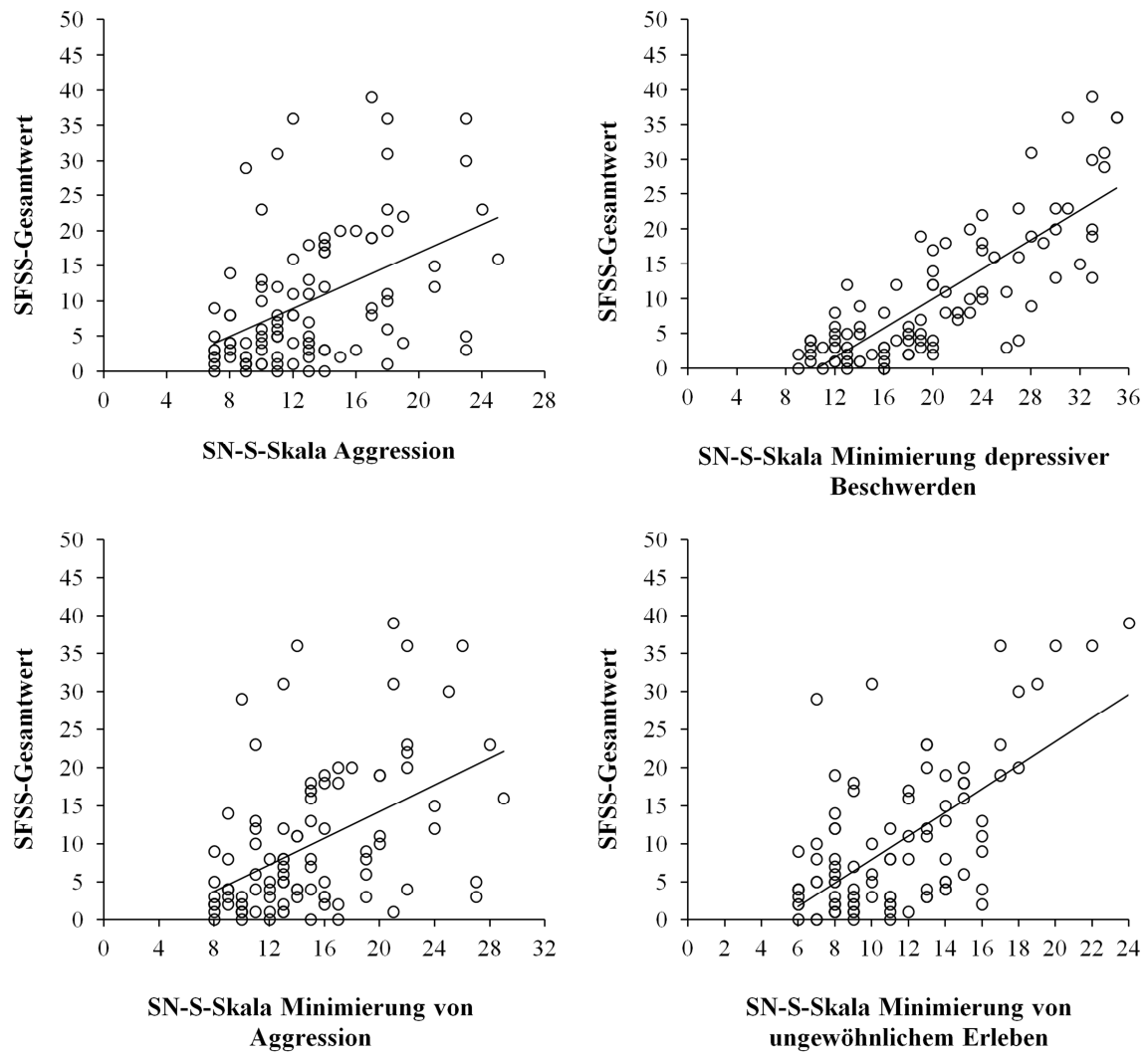


Abbildung 5. Zusammenhang zwischen dem SFSS-Gesamtwert und den SN-S-Skalen Aggression, Minimierung depressiver Beschwerden, Minimierung von Aggression und Minimierung von ungewöhnlichem Erleben

Tabelle 29

Deskriptive Statistiken des SN-S-Gesamtwerts und der Subskalen in der Gesamt- und den Substichproben (Fortsetzung S. 148)

	GS	SF	RP	ZB	GN	KG	Fake-KG
SN-S-GW							
<i>M (SD)</i>	63.4 (15.4)	71.5 (21.3)	60.1 (13.3)	77.6 (19.8)	52.1 (10.4)	66.8 (11.9)	51.8 (7.1)
<i>Mdn</i>	61.0	67.5	60.0	82.0	50.5	67.0	51.5
<i>Min-Max</i>	34-117	34-117	40-93	41-113	43-76	42-106	35-69
SN-S-SE							
<i>M (SD)</i>	10.1 (2.5)	11.0 (2.7)	10.7 (1.9)	10.9 (2.6)	8.5 (1.8)	10.6 (2.4)	8.3 (1.7)
<i>Mdn</i>	10.0	11.0	11.0	11.0	8.5	11.0	9.0
<i>Min-Max</i>	4-16	4-16	6-16	5-16	6-11	5-15	5-13
SN-S-AS							
<i>M (SD)</i>	14.5 (4.6)	16.7 (5.5)	13.5 (4.6)	19.5 (2.6)	11.6 (3.0)	15.2 (3.9)	11.4 (2.6)
<i>Mdn</i>	14.0	15.5	13.0	11.0	11.5	15.0	11.0
<i>Min-Max</i>	7-28	7-28	7-25	5-16	8-17	8-27	7-21
SN-S-ZW							
<i>M (SD)</i>	6.7 (2.3)	7.9 (3.1)	5.9 (2.0)	8.6 (2.9)	6.0 (2.0)	7.0 (2.1)	5.4 (1.3)
<i>Mdn</i>	6.0	7.5	5.0	9.0	5.5	7.0	5.0
<i>Min-Max</i>	3-14	4-14	4-11	4-14	4-10	3-13	4-9
SN-S-DE							
<i>M (SD)</i>	9.3 (2.8)	10.1 (3.8)	9.2 (2.2)	11.0 (3.9)	8.4 (2.6)	9.7 (2.5)	7.8 (1.8)
<i>Mdn</i>	9.0	10.0	9.0	11.0	8.0	10.0	8.0
<i>Min-Max</i>	5-20	5-18	5-14	5-20	6-14	5-17	5-13

Fortsetzung Tabelle 29	GS	SF	RP	ZB	GN	KG	Fake- KG
SN-S-PS							
<i>M (SD)</i>	8.8 (3.0)	9.8 (4.7)	7.7 (2.3)	10.4 (3.9)	7.5 (1.9)	9.4 (2.6)	7.6 (1.6)
<i>Mdn</i>	8.0	8.0	7.0	10.0	6.5	9.0	7.0
<i>Min-Max</i>	5-24	6-24	6-18	6-20	6-11	5-20	6-12
SN-S-AGG							
<i>M (SD)</i>	12.2 (3.6)	13.9 (5.0)	11.5 (3.4)	14.4 (4.6)	9.1 (1.6)	13.2 (3.0)	9.8 (2.1)
<i>Mdn</i>	12.0	13.0	11.0	14.0	9.0	13.0	10.0
<i>Min-Max</i>	6-25	7-25	7-23	7-24	7-12	8-22	6-16
SN-S-MDe							
<i>M (SD)</i>	17.8 (6.1)	20.3 (7.2)	16.7 (5.6)	25.4 (7.1)	14.1 (3.8)	18.6 (5.1)	13.6 (3.0)
<i>Mdn</i>	12.0	20.0	16.0	24.0	12.5	18.0	10.0
<i>Min-Max</i>	9-35	9-35	9-33	10-35	11-22	10-33	6-16
SN-S-MA							
<i>M (SD)</i>	14.2 (4.3)	15.9 (5.7)	13.3 (4.1)	16.7 (5.4)	10.8 (1.5)	15.4 (3.6)	11.2 (2.2)
<i>Mdn</i>	13.0	14.0	13.0	16.0	10.0	15.0	11.0
<i>Min-Max</i>	8-29	8-29	8-27	8-28	9-13	9-26	8-18
SN-S-MUE							
<i>M (SD)</i>	10.5 (3.4)	12.2 (4.0)	9.7 (2.8)	12.2 (4.7)	9.4 (3.3)	11.1 (3.2)	8.6 (2.0)
<i>Mdn</i>	10.0	12.0	9.0	11.0	8.0	10.5	8.0
<i>Min-Max</i>	6-24	6-22	6-16	6-24	6-14	6-23	6-15

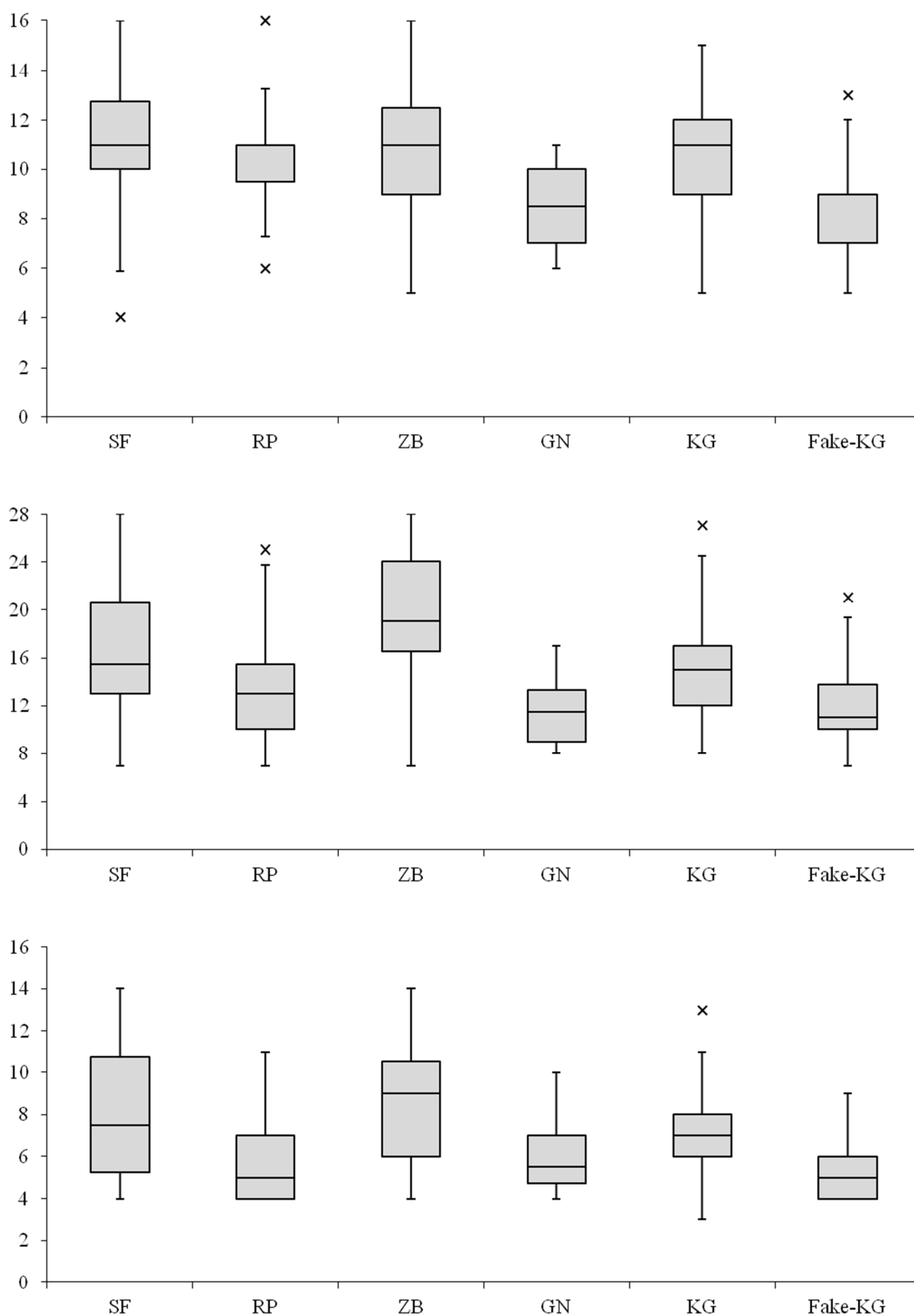


Abbildung 6. Verteilung der Werte auf den SN-S-Skalen Soziale Erwünschtheit (4 Items), Affektive Symptome (7 Items) und Zwang (4 Items) in den Substichproben

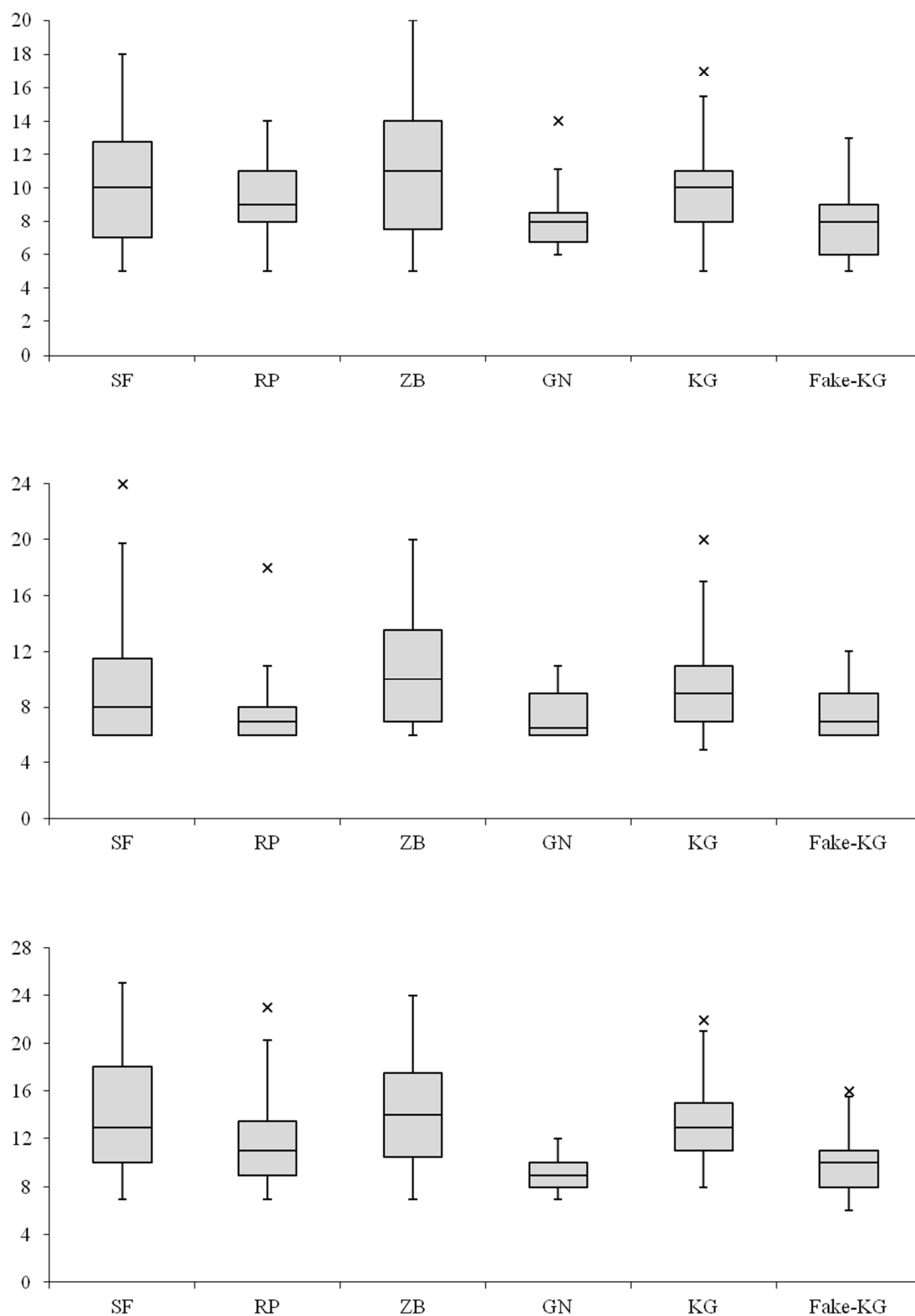


Abbildung 7. Verteilung der Werte auf den SN-S-Skalen Dissoziatives Erleben (5 Items), Psychotische Symptome (6 Items) und Aggression (7 Items) in den Substichproben

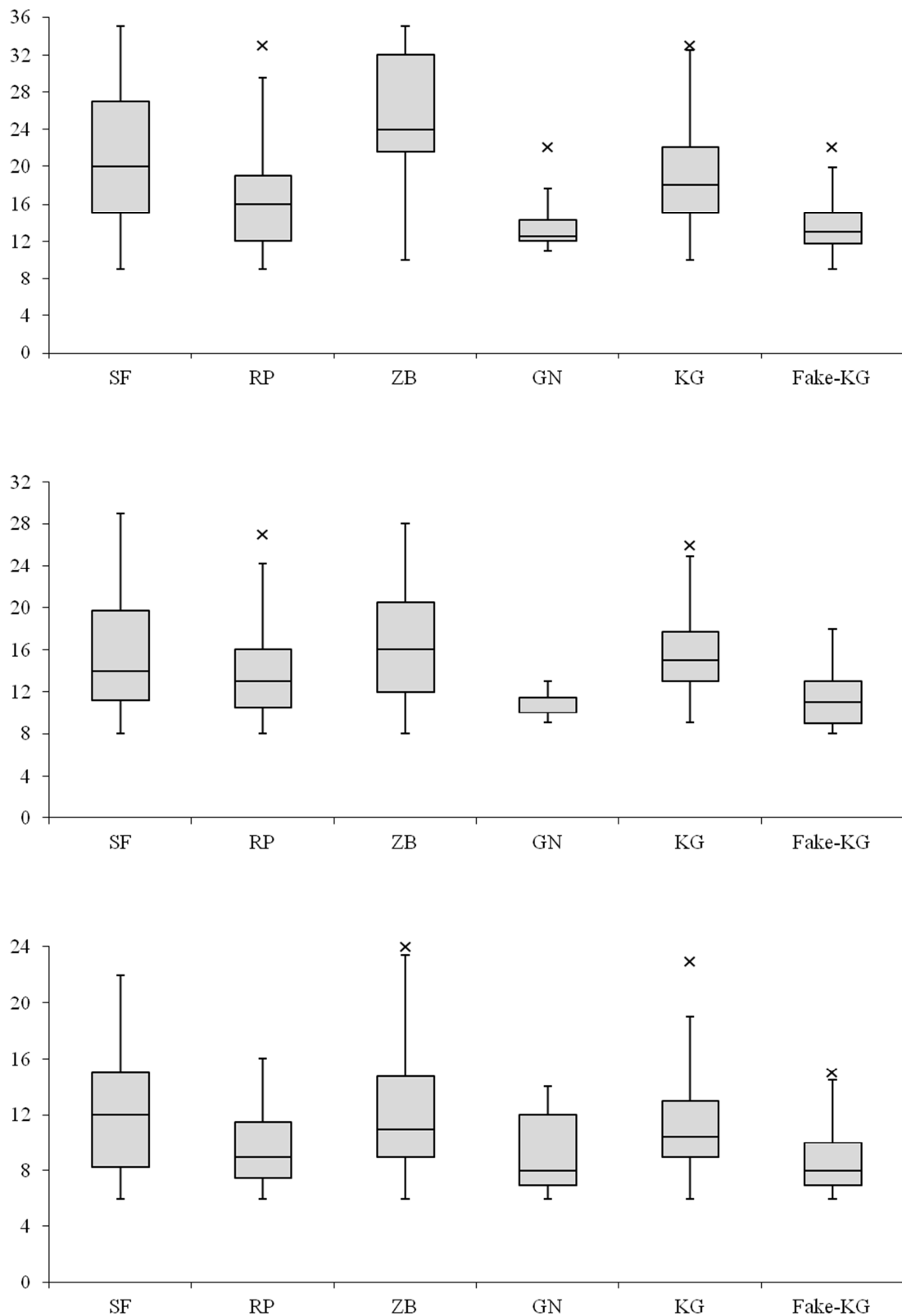


Abbildung 8. Verteilung der Werte auf den SN-S-Skalen Minimierung depressiver Beschwerden (9 Items), Minimierung von Aggression (8 Items) und Minimierung von ungewöhnlichem Erleben (6 Items) in den Substichproben

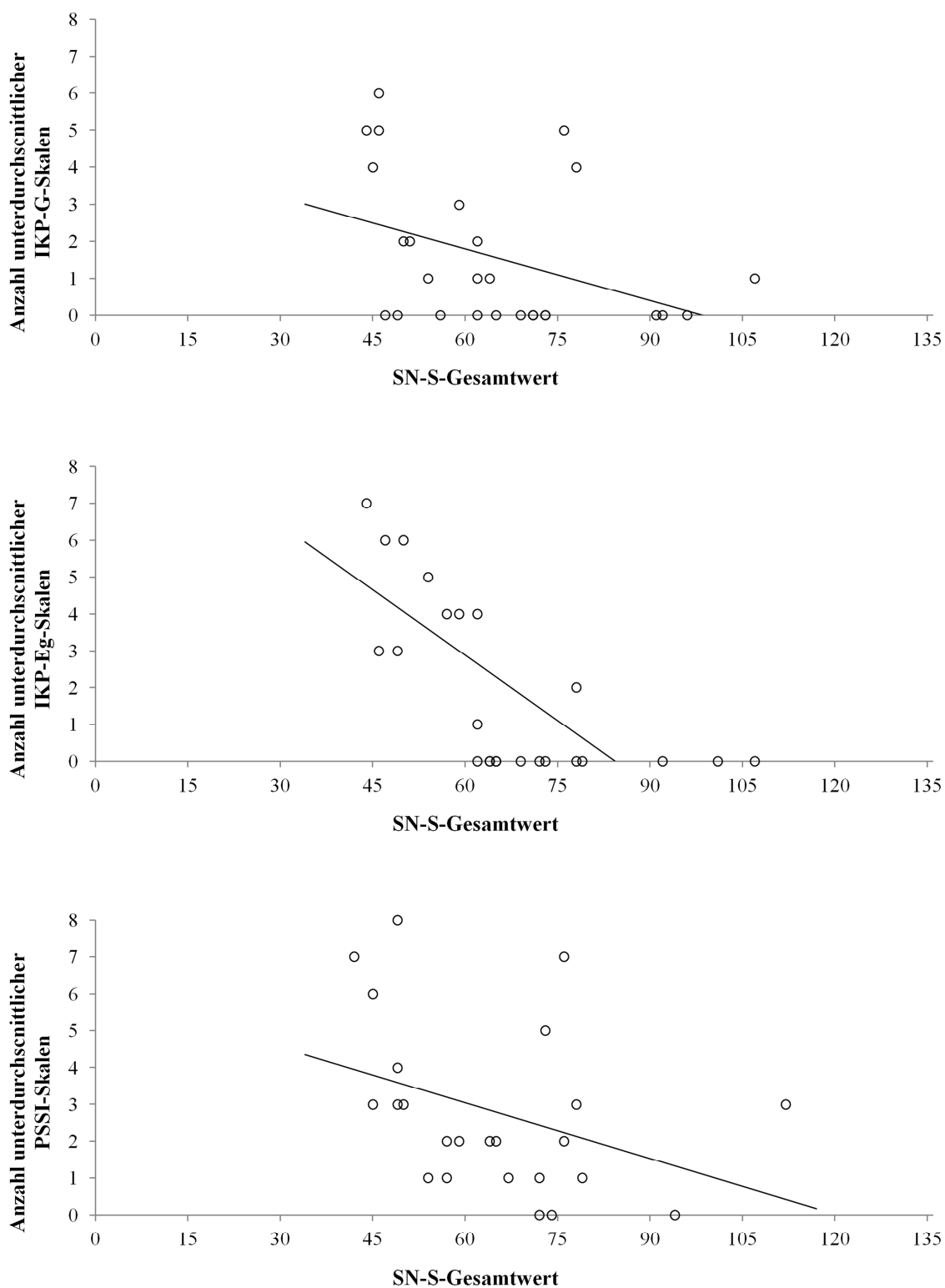


Abbildung 9. Zusammenhang zwischen dem SN-S-Gesamtwert und der Anzahl unterdurchschnittlicher Skalen in IKP-G, IKP-Eg und PSSI

Danksagung

Mein besonderer Dank gilt meinem Betreuer Prof. Dr. Norbert Nedopil für die Unterstützung bei dieser Arbeit und insbesondere für die fundierte Ausbildung im Bereich der forensischen Psychiatrie, die ich in seiner Abteilung genießen durfte.

Auch meiner lieben Kollegin Frau Dr. Elena Yundina möchte ich danken für die hilfreichen Anregungen bei der Vorbereitung und Durchführung der Datenerhebung und bei dem Verfassen der Arbeit.

Ganz herzlich bedanke ich mich bei Michael Schneider und meiner Familie, die mir die ganze Zeit über mit viel Geduld zur Seite standen, vor allem natürlich Sabrina Appelshäuser, die wie immer aufmerksam Korrektur gelesen hat.

Eidesstattliche Versicherung

Name, Vorname

Ich erkläre hiermit an Eides statt,
dass ich die vorliegende Dissertation mit dem Thema „Diagnostik positiver Antwortverzerrungen bei forensischen Begutachtungen. Untersuchung der Validität und Brauchbarkeit der Supernormalität-Skala (SN-S)“ selbstständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum

Unterschrift